University of New South Wales
School of Economics

Honours Thesis

Existential Risk and Pandemic Preparedness Spending

Author:
Thomas Houlden
Student ID: 5161531

Supervisor:
Dr. Gigi Foster

B. Ec (Economics) (Honours)
AND
B. A (Philosophy)

Submitted 19 November 2021

Declaration

I hereby declare that the content of this thesis is my own work and that, to the best of my knowledge, it contains no material that has been published or written by another person or persons, except where due acknowledgement has been made. This thesis has not been submitted for award of any other degree or diploma at the University of New South Wales or any other educational institution.

Thomas Houlden
13th May, 2022

Acknowledgements

I would like to thank my supervisor, Dr Gigi Foster, for her guidance, knowledge, and support over a number of years during my studies at UNSW. I am eternally grateful for the many opportunities and experiences that have been afforded to me by Dr Foster. I would also like to acknowledge Dr Sang-Wook Cho, whose generosity with his time and expertise has provided an invaluable learning experience.

I would also like to express my gratitude to Annabel Burnett, Josh Levy and Dr Gabriele Gratton for their many helpful comments on drafts of this thesis; and to the UNSW Economics Honours cohort of 2021 for their warmth and curiosity. All of these passionate people have made this year as enjoyable and rewarding as I could have possibly hoped.

I would also like to thank Aram Perez, not only for his very detailed remarks on the mathematical components of this thesis, but also for our many formative conversations which have inevitably shaped my thinking about many of the concepts raised and explored in this thesis.

Finally, I am grateful to the Effective Altruism community which has fostered a dialogue around existential risk and catastrophe and has been a constant source of inspiration and motivation throughout the writing of this thesis. In particular, I am thankful to David Janků and Effective Thesis for their comments and guidance in selecting my thesis topic.

Contents

List of Figures

List of Tables

Abstract

Existing cost-benefit analyses of interventions to mitigate pandemic risk have tended to assume that pandemics are not a threat to the continued existence of our species. This thesis takes a novel approach to assessing the consequences of this assumption by evaluating society's willingness to pay (WTP) for the mitigation of risk from pandemics while varying estimates of the level of existential risk with the application of an endogenous discount factor. The analysis demonstrates that, after including existential risk as a possible consequence of pandemics, a specific pandemic risk mitigation intervention only has to be, at the most, half as effective as previously estimated to break even. This analysis also suggests that the total WTP to mitigate all pandemic threat can increase by between 6 and 10% under 'best-guess' calibrations. Importantly, the results presented are highly dependent on the calibrations of key variables including aversion to intergenerational inequality, the level of existential risk that can be mitigated by an intervention, the level of population growth, and the method of aggregating individual utility. Since these variables are largely calibrated by subjective estimation, this analysis represents a numerical experiment rather than offering precise estimations.

Chapter 1
Introduction

Existential threats are the class of catastrophe that could result in human extinction, or permanently and dramatically reduced human welfare (Bostrom, 2002). Given the potential for hundreds of thousands or even millions of years of future human generations and the possibility of events that might snuff out this potential, interventions which decrease the risk of existential threats by even a meagre amount could, *prima facie*, be tremendously valuable.

While we have no empirical reference for human extinction, we have had glimpses of events that have been damaging or threatening enough so that the possibility of a more extreme, species-threatening version is not too far-fetched. To name a just few examples, between the years of 1347 and 1353 the Black Death killed between one-quarter and one-third of the global population (Benedictow, 2004; Ziegler, 1969); 65 million years ago a ten-kilometre wide asteroid was responsible for the wiping out of three-quarters of all species on earth, concluding the age of the dinosaurs (Schulte et al., 2010); and the 20th century saw 32 documented nuclear accidents involving U.S. weapons, some of which could have easily resulted in millions of deaths (US Department of Defence, 1981).

Of course, presently the COVID-19 pandemic is loosening its grip on our species – an event which, at the time of writing, has infected over 200 million people, of whom over 4.7 million have been confirmed dead (Roser et al., 2020).[1] This pandemic has an infection fatality rate (IFR) of, at the very most, 1.5%, meaning that even if the virus infected every single human we could only expect to lose, at most, a very small fraction of the global population (Ioannidis, 2020). In many analyses of the value of spending on mitigating pandemic risk, there is an implicit assumption that the pandemics we face will be somewhat like the COVID-19 pandemic – resulting in the deaths of some proportion of the global population and perhaps resulting in some loss of GDP from which we will inevitably recover. However, as will be discussed in Chapter 2, there are good reasons to expect that there is at least some probability, even if this probability is very, very small, that we could encounter a pandemic from which we do not recover. If so, these existing analyses will underestimate the true value of spending on pandemic mitigation.

[1] Though the true death toll, as measured by excess mortality, is expected to be between 1.3 and 2.2 times higher (Sanmarchi et al., 2021).

The focus of this thesis is to develop and apply a framework for evaluating resource allocation to mitigate the risk of existential threats (hereafter existential risk), with a particular focus on pandemics. Specifically, I hope to answer two questions. Firstly, under what calibrations of key variables (including preferences related to intergenerational inequality, probability of existential collapse, and population growth) and classes of social welfare function should a social planner be willing to dedicate significant resources to existential risk mitigation? Secondly, how do existential considerations alter the value of pandemic mitigation activities?

This thesis provides a more rigorous evaluation of the consequences of existential risk for spending decisions than what has been offered in a previous analysis focused on existential risk and biosecurity. For example, Millet and Snyder-Beattie (2017) offer an estimation of the cost-effectiveness of interventions to mitigate existential risk from pandemics and other biothreats; however, the authors do not consider the dynamic effects of a pandemic shock and provide an analysis that is inconsistent with more recent literature on the evaluation of existential threats.[2] Instead, I take an endogenous discounting approach to estimate the willingness-to-pay (WTP) for existential threat mitigation that has, to my knowledge, not been applied in the context of catastrophic impacts of pandemics. This approach has recently been applied to catastrophic climate change in Méjean et al. (2020), to estimate the optimal greenhouse gas (GHG) emission path depending on key variables such as aversion to intergenerational inequality, the type of social welfare function applied, and the marginal impact on existential risk GHGs. In this thesis I expand on the work of Méjean et al. (2020), not only through a novel application of the endogenous discounting approach but also by developing a model that can be applied more broadly and demonstrating some theoretical implications of this model.

The results presented in Chapters 3, 4, and 5 are consistent with many of the general findings of Méjean et al. (2020). For example, I demonstrate that WTP for existential risk mitigation diminishes strongly with greater aversion to intergenerational inequality, and similarly with social welfare functions where the marginal gains in aggregated individual utility diminish with population size (Chapter 3). I then apply the endogenous discount approach to evaluate pandemic preparedness spending, showing that a specific intervention to mitigate pandemic risk is up to twice as cost-effective than initially thought if it can mitigate some proportion of total existential risk; though this effect is highly dependent on

[2]Specifically, Millet and Snyder-Beattie (2017) do not evaluate the impact of existential risk with a social welfare function with an endogenous discount factor. Such an approach does not rely on the same strong assumptions regarding the expected number of future generations and is grounded in standard economic variables (such as growth, preferences related to inequality, and, of course, the discount rate).

how much risk is mitigated by the intervention in question (Chapter 4). I also demonstrate a similar effect when the endogenous discount model is applied in a dynamic environment, and under the most likely calibration of key variables in this model we can see a 6-10% increase in WTP after including existential considerations (Chapter 5).

The motivation for this focus on pandemics is twofold. Firstly, there has been a recent explosion of literature on the economics of pandemics with a particular focus on the impact of the COVID-19 pandemic. This literature has ignored the possibility of a pandemic representing an existential threat. Secondly, biosecurity has recently been ranked as the most neglected risk to future human welfare in a recent report on extreme risk, due to the scale of the threat and the currently limited measures put in place to mitigate biorisk (The Centre for Long-Term Resilience, 2021). I hope this thesis makes a modest contribution to improving our (currently limited) understanding of the scale of the threat from future pandemics.

Chapter 2
Background

2.1 Existential Threats

2.1.1 The Possibility of an Existential Threat

Humanity has faced tremendously damaging catastrophes, though the fact that we are alive today means that our species has, so far, avoided an existential collapse. One possible explanation for the absence of this collapse is that humanity does not actually face existential threats – that the kinds of catastrophes we have seen do not have the power to extinguish humanity's potential entirely. I briefly note two arguments against this claim.

The first argument to resist this claim is that extinction is hardly a rare phenomenon. The scientific consensus is that the natural extinction rate (outside of human intervention) for all species on earth is 1 species in 10,000 per 100 years. However, the current extinction rate is placed somewhere in the range of 8 to 100 times greater than this depending on the animal group (Ceballos et al., 2015). For mammals, the average species lifespan is 2 million years (Avise, Walker, and Johns, 1998), and our nearest ancestor, *homo erectus*, died out after 1.6 million years of existence (Antón, 2003). Comparatively, *homo sapiens* have only existed for 200,000 to 300,000 years (Galway-Witham and Stringer, 2018). If our ancestors serve as any guide then our current survival should not be taken as evidence of our resistance to existential threats since by comparison we are still in our infancy.

The second reason is that taking our current existence as evidence of resistance to existential collapse assumes a strong degree of past-future symmetry – that the threats in the past mirror the threats we face today, and will face in the future. This assumption seems exceedingly difficult to justify given the new and evolving threats that we face today. It is fitting that our current epoch, the Anthropocene, is defined by our newfound capacity to transform our environment, both destructively and constructively.[1] Given this new era marked by humanity's potency, the past–future symmetry assumption seems to fall very short of being able to offer us any reasonable estimate of the likelihood of future existential threats (Ćirković, 2008). Without history offering a reliable yardstick, out of necessity we have to turn to *ex*

[1] It seems relevant that the Anthropocene is dated by many to the explosion of the first atomic bomb on July 16th, 1945 (Oreskes et al., 2015).

ante estimates of the probabilities of various existential threats. Of course, these 'best-guess' estimates are necessarily more subjective,[2] but at least they are sensitive to the risks unique to the Anthropocene.

One may resist this second reason by pointing out that technological improvement can also serve to increase safety. For example, COVID-19 vaccines were developed faster than any vaccine in history, largely due to innovations in vaccine development (Ball, 2020).[3] Despite these 'safety' technologies, researchers in catastrophic risks tend to be most concerned about new technological developments which might result in an existential collapse before humanity can respond effectively to that new threat.[4] For example, prior to the first nuclear explosion, it was a credible scientific position to hold that such an explosion would be powerful enough to ignite the hydrogen in the atmosphere and oceans, eradicating humanity and potentially all complex life (Ord, 2020). So while it is true that technological improvements can mitigate existential risk, many of the risks to future human existence arise from new technologies for which it is difficult to anticipate the necessary precautionary measures.

The most defining catalogue of estimates of the probability of various existential threats – the risk landscape – was recently proposed by Ord (2020). Ord's appraisal is that over the next century there is a 1 in 10,000 chance of extinction from natural events (asteroid impact, super-volcanic eruption, or stellar explosion); a 1 in 1,000 chance from each of nuclear war, climate change and other environmental damage; a 1 in 10,000 chance from naturally arising pandemics and a 1 in 30 chance from engineered pandemics; a 1 in 10 chance from 'unaligned' artificial intelligence (an artificial intelligence which pursues ends not in accordance with human values); a 1 in 50 chance from 'other anthropogenic risks' (such as nanotechnology, or back contamination from space exploration); and a chance 1 in 30 from unforeseen events. The result is a rough estimate of a 1 in 6 chance of existential catastrophe in the next century (Ord, 2020, pg. 167). Ord stresses the imprecise nature of these estimates and suggests interpreting them as representative of orders of magnitude rather than scientific or objective conclusions.

Other works have tended to recognise similar events within the class of existential threats though have offered differing probability estimates (though roughly similar in orders of magnitude for the various threats). For example, a survey of

[2] In this context, by 'subjective' I mean derived with some significant degree of guesswork and intuition rather than just empirical evidence. Though this alone should not deter us from applying these estimates, since any approach to predicting existential risks is necessarily subjective in the same way.

[3] Another example of how technology can be used to mitigate existential threats is the present use of telescopic surveys to detect near-earth asteroids (NEAs) which could potentially result in catastrophic consequences for life on earth.

[4] See the 'vulnerable world hypothesis' suggested in Bostrom (2019).

attendees at the 2008 Global Catastrophic Risk Conference revealed various median estimates for existential threats by 2100: from both molecular nanotechnology and unaligned artificial intelligence, 1 in 20; an engineered pandemic, 1 in 50, and a naturally arising pandemic, 1 in 2000; nuclear war, 1 in 100, and nuclear terrorism, roughly 1 in 3000. The result is a median estimate of a 1 in 5 chance of extinction by 2100 (Sandberg and Bostrom, 2008).[5]

Given the subjective nature of these estimates, one may reasonably disagree with these characterisations of the risk landscape. Therefore, in the economic analysis of existential risk offered in this thesis I consider several calibrations for existential risk, spanning several orders of magnitude. Consequently, the approach of this thesis should be interpreted as a numerical experiment rather than a scientific or objective analysis.

2.1.2 Pandemics as an Existential Threat

Humanity is yet to encounter a pandemic that has seriously threatened the very existence of humankind, though there have been examples of severe pandemics which have resulted in the death of a significant portion of the global population. The Black Death has already been mentioned, which resulted in deaths estimated to be in the range of a quarter to a third of the global population. Another severe pandemic, the Plague of Justinian, circa 541-750, has been estimated to have killed between 30% and 60% of the population of the Mediterranean at the time (Allen, 1979; Harper, 2016), though recent work has suggested the lower figure to be more likely (Mordechai and Eisenburg, 2019; Mordechai et al., 2019). A more recent example of a particularly catastrophic pandemic is the Spanish Flu, which resulted in 50 million deaths after a 1918 outbreak – roughly 1% of the global population at the time. Since these are the worst of the pandemics, this seems to provide some solace – as put in Bostrom (2013), '[h]umanity has survived what we might call natural existential risks for hundreds of thousands of years; thus it is *prima facie* unlikely that any of them will do us in within the next hundred' (pg. 15).

However, recent work on global pandemics and catastrophic risk has suggested that, in fact, we ought to recognise natural pandemics as an existential threat, even if this threat is relatively small. One reason for this elevated concern, despite the historical cases of pandemics not being anywhere near this severe, is the increased intensity of animal husbandry, increased antimicrobial resistance, increase human population density, and climate change which are all suggested to increase the rate of emerging infectious diseases (EIDs) (Jones et al., 2008). It is difficult to tell whether this increase is sufficient to place naturally occurring pandemics into the class of

[5] Further arguments for these candidate threats can be found in Posner (2004), Bostrom (2002), Ng (2016), Chichilnisky and Eisenberger (2010), Wilson (2013), Taleb et al. (2014a), and Leigh (2021).

existential threats, though given the concern from experts in catastrophic risk noted in the previous section, it seems we should calibrate the possibility of existential collapse from naturally occurring pandemics to some non-negligible figure, even if that figure does seem very low – I consider a 1 in 1,000,000 annual chance as a baseline calibration in the analysis later in this thesis following Ord (2020).

When it comes to existential risk from pandemic threat, risk research has tended to attribute the bulk of the threat to engineered or lab-leaked pandemics. The threat from these materials has increased dramatically over the last half-century, with significant advances in biotechnology and particularly with the accessibility and sophistication of DNA sequencing tools.[6] With this technological improvement we have seen a democratisation of biotechnology, but as noted by Ord, 'democratisation also means proliferation' (pg. 134) and consequently an increase in the probability of a malicious actor having access to the skills and technology necessary to craft deadly biological agents. This use of biotechnology to construct deadly pandemics is not new: Ord notes 15 countries who have employed biotechnology for these ends during the 20th century which saw the attempted weaponisation of diseases such as smallpox, anthrax and tularemia, and Leigh (2021) lists several examples of malicious, non-state actors who have employed bioweaponry to cause death and destruction.

The main concern about the possibility of an existential threat from pandemics comes from a combination of two facts I have presented here: we have observed very severe pandemics in the past which have killed a lot of people, and the proliferation and improvement of gene editing tools mean that bioweaponry will become more accessible for malicious actors and more deadly in their hands. In light of these two facts, it seems likely that actors with malicious intent and some knowledge of biological structures and human anatomy could eventually deliver a pandemic significantly worse than those which nature has stumbled upon so far.

2.2 The Economics of Catastrophes

2.2.1 Uncertainty

One of the major debates in the economics literature on catastrophes is the suitability of standard economic tools (such as CBA and expected utility theory) to evaluate actions to mitigate catastrophic risk. One of the defining arguments in this literature is the 'dismal theorem' in Weitzman (2009), which suggests that evaluation of catastrophe rests on 'deep structural uncertainty'. Economic models

[6] For example, see Nouri and Chyba (2008), National Research Council (2006), or Tucker and Zilinskas (2006) for a discussion of the threat from improving biotechnology, and Klotz (2019) for a discussion on the threat from lab-leaks.

of such catastrophes are subject to a 'best-guess' parameterisation, and therefore are highly sensitive to highly subjective estimates. Consequently, these economic models tend to be an exercise in the art of developing convincing estimates (which cannot be empirically verified), rather than a science of assessing costs and benefits.

Economists have debated the force of the dismal theorem. For example, both Millner (2013) and Nordhaus (2011) argue that the dismal theorem only holds under a limited set of conditions (these conditions relate to the level of risk aversion, the size of the fat-tail,[7] and the inability of society to learn). Furthermore, Nordhaus argues there is no clear rule baked into the dismal theorem about how much uncertainty is too much (this is particularly important since some uncertainty is common in many broadly useful economic models) and that this should motivate researchers to reduce uncertainty around these events so that standard economic techniques can be applied more effectively.

Uncertainty is clearly an obstacle for economists working on catastrophic events, though the applied literature has developed norms around dealing with this level of uncertainty. One approach is to simply make conservative assumptions around some event and understand whether under these conservative assumptions some particular course of action is still advisable under CBA. A second, numerical experiment approach is to explore several different parameterisations of uncertain variables – the result being an 'if/then' recommendation where *if* we assume a particular parameterisation, *then* this implies some optimal course of action. In the literature these approaches tend to be applied together, where a researcher uses an array of parameterisations, beginning with the most conservative and then demonstrating how the resulting recommendation changes as the parameterisation becomes less and less conservative.[8] A third approach is to develop estimates for both the costs and benefits of an intervention and then calculate the 'break-even' change in probability – the change in probability such that the costs of the intervention are exactly equal to the expected benefits from the intervention. For example, if a particular intervention to reduce the likelihood of Event X costs $100 million dollars to implement, and should Event X transpire it would cause $1 billion worth of damages, then the break-even change in probability is the costs divided by the damages (here 10%).[9]

These methods enable analysts to come up with a defined figure for the rate of return, cost per life saved, break-even probability or some other metric;

[7] In a 'fat-tail' distribution, the probability of a very extreme outcomes is relatively higher than a tailed distribution.

[8] For example, Méjean et al. (2020) evaluates optimal climate policy in this way; World Bank (2012) and Martin and Pindyck (2021) develop estimates related to pandemic policy in this way.

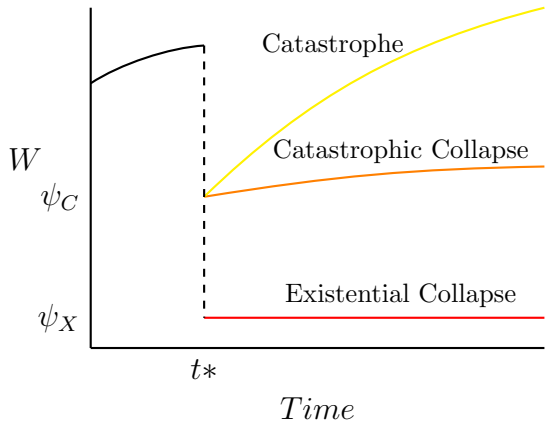
[9] For example, Dobson et al. (2020) conducts a CBA on an intervention to reduce pandemic probability in this way, which I consider more carefully in Chapter 4.

though, they are unable to resolve the uncertainty problem, only push through it. Furthermore, any application of these results requires the decision-maker to confront these uncertainties anyway. In the above example of Event X, the question is whether the planner thinks the intervention reduces the likelihood of Event X by more or less than 10%. If Event X represents any of the existential threats discussed in this thesis, then any answer to this question is still a subjective best guess, though the problem has become somewhat easier than trying to nail down a precise estimate of the effectiveness of the intervention.

2.2.2 Permanent Collapse

The literature on catastrophes defines collapse as an unrecoverable decrease in welfare, whereas a mere catastrophe can be defined as a momentary shock to welfare that is recovered. Here I further separate a collapse into a catastrophic collapse, where welfare is permanently reduced (inducing a 'regime shift'), and an existential collapse where welfare is drastically reduced to ψ_X, this is close to zero.[10] Representative examples of these three categories are presented in Figure 2.1 after a catastrophic event at $t*$.

Figure 2.1: Long-run Welfare Consequences of Catastrophes



Compared to a catastrophe, an existential collapse reduces welfare more dramatically and permanently; compared to a catastrophic collapse, an existential collapse reduces welfare more dramatically.

There has been some analysis of both existential and catastrophic collapses. For example, Bommier, Lanz, and Zuber (2015) define the optimal growth path under the possibility of catastrophic collapse induced by carbon emissions. By contrast, both Cropper (1976) and Clarke and Reed (1994) consider optimal growth paths under the possibility of an existential threat from environmental collapse which

[10] To my knowledge no literature has defined ψ_X precisely.

permanently reduces social welfare to zero. This work has shown, unsurprisingly, that if the hazard rate increases with emission (or pollution) stock then this lowers optimal equilibrium emissions.

2.2.3 Existential Risk and Discounting Utility

While there has been little economic research explicitly focused on extinction risk, there have been allusions to such risk in the established literature on discounting. In any project considering the value of altering the consumption or utility (or both) of future generations, economists are familiar with applying some level of discount – though the scale of and motivation for discounting is quite divisive. The usual application of future utility discounting is in a welfare function:

$$W_i = \sum_{t=0}^{T} \frac{u_i(c_{ti})}{(1+\delta)^t}$$

where welfare for individual or household i is defined by the finite utility stream, u, which is a function of consumption, c, in each period, t. This utility is discounted by a factor δ, where $0 < \delta < 1$, which has an exponent of t, meaning if consumption were constant, the value from each period of future utility decreases as t increases. This discount on utility (or pure time preference) is not the same as the discount on future consumption, which factors in the growth rate and marginal utility of consumption as well as pure time preference.[11]

The specification for δ found in the literature tends to be either in the order of 1-2% or 0-0.1%, representing two different interpretations of pure time preference. While this difference may appear to be small, over long periods of time this difference is highly consequential. For example, at $\delta = 1.5\%$ the present value of 100 units of utility in 100 years time is only 22.5 compared to 90.5 at $\delta = 0.1\%$.

The first interpretation of δ is a descriptive (positive) one – discounting is the result of preferences that are revealed through market activity. The second interpretation is a prescriptive (normative) one – discounting utility reflects social values towards future lives and these values are not expressed in market activity.[12]

The descriptive camp derives a value for δ by, first, looking to the long term interest rate, r, and then working backwards from the formula in Ramsey (1928):

$$r = \delta + \eta g$$

[11] Although, as we will see in Section 3.1 one might consider inequality aversion either across generations or within generations instead of marginal utility of consumption; see Greeves (2017).

[12] It was disagreement between these two camps that drove the controversy surrounding the Stern Report on Climate Change (2007), which took a normative approach to discounting. Criticisms from Weitzman (2007) and Nordhaus (2008) proposed applying a pure time preference derived descriptively, which dramatically changes the resulting policy recommendation.

where g is the growth rate of the economy, and η is the marginal utility of consumption.[13] For example, Weitzman (2007) takes $g = 0.02$, $\eta = 2$, and $r = 0.06$, implying $\delta = 0.02$.[14]

The prescriptive camp derives a value for δ with philosophical argument. The standard line of argument is that that there is no reason to weight the preferences of those living more highly than future generations.[15] Even if this is true, we might still want to discount later enjoyments for uncertainty reasons – if I am not totally confident that I will receive some future benefit (perhaps because there is some daily risk of death which could prevent me from accessing some future good), then I would reasonably discount the present value of this enjoyment even if I weight future utility equally to present utility. Following this logic, researchers in the descriptive camp have advocated for a discount on pure utility as an 'extinction discount', whereby the annual discount on pure utility is allowed to represent the probability that future people will be alive to enjoy future benefits.[16] For example, in Stern (2007) the pure time discount is set at 0.1%, corresponding to what Stern places as a 90% chance of surviving the next century. Ng (2016) has argued that Stern's value is set too high, advocating for a value 'at least 10 times smaller' at $\delta < 0.01\%$ (corresponding to the probability of at least 99% of surviving the next century). Appendix C provides a comprehensive list of the different calibrations for δ suggested in previous literature; these calibrations tend to favour calibrations of risk in the order of magnitude suggested by Stern rather than Ng. Given the typical calibrations for δ in the literature and the research described in section 2.1 related to the new risks of today and the future, I suspect Ng's calibration is too optimistic.

This interpretation of δ as extinction probability opens new possibilities for analysis under the possibility of extinction risk. In particular, allowing δ to be endogenous in a social welfare function for analysis of catastrophes enables an assessment of the value associated with a change in the expected date of human expiration. Though this approach also has some limitations, namely that it assumes that an existential catastrophe happens immediately so that welfare drops suddenly rather than being gradual decrease. This possibility is considered more carefully in Chapter 6. Recently, Méjean et al. (2020) have taken this approach to evaluate optimal climate policy given the (narrow) possibility of extreme consequences of climatic change. This model provides a framework to trade-off between costs

[13] Though η has a number of interpretations which will be discussed in Chapter 3.

[14] See also Nordhaus (2008) and Nordhaus (2006).

[15] For example, Ramsey (1928) writes: 'we do not discount later enjoyments in comparison with earlier ones, a practice which is ethically indefensible and arises merely from the weakness of the imagination'.

[16] This argument can be found, for example, in Mirrlees (1967), Dasgupta and Heal (1979), Stern (2007), Chichilnisky, Hammond, and Stern (2020), Ord (2020), Ng (2016), and Méjean et al. (2020).

from reduced productivity from stronger climate action and benefits from decreased expected future damages and a decreased extinction probability. The findings from this analysis suggest that existential considerations can have a significant effect on the optimal climate policy recommendation under certain assumptions about the level of intergenerational inequality aversion, and the scale of existential risk from climate change.

Given that many economists choose to interpret δ in existential terms, it is surprising that the endogenous discounting approach has received little attention in deriving optimal policies or in CBA for interventions to reduce potentially catastrophic events. One difficulty with the application of an endogenous discount model to pandemics, relative to its application in catastrophic climate change by Méjean et al. (2020), is that there do not exist sophisticated models which can capture the relationship between capital allocation resulting in changes in key variables including consumption, education, technology, growth, and of course pandemic probability, to assess optimal allocation.[17] Such models cannot be easily developed for the effects of capital allocation to pandemic mitigation since the response of pandemic risk to such an allocation is unobservable. Therefore, I take a novel approach to understanding the impact of existential risk on spending decisions for pandemic mitigation by assessing WTP, which can provide valid results in the absence of macroeconomic models to assess effects of capital allocation on the wider economy, which would be required to assess the optimal investment. However, this approach does mean that there are limitations in the application of this research to questions of public policy, namely that WTP offers little practical guidance in terms of how much should be dedicated to some policy or intervention.[18] Despite this, the WTP approach is appropriate to provide a sense of scale of the importance of existential considerations when spending on mitigating catastrophic threats such as pandemics.

I begin my analysis by exploring results for WTP for existential risk mitigation by applying a social welfare function with an endogenous discount factor. Since evaluating WTP is a novel application of this endogenous discount model, first I demonstrate how WTP responds to changes in key variables that are demonstrated by Méjean et al. (2020) to have an effect on the optimal policy after including existential risk in their model. Next, I apply this model to pandemics to, first, understand how this might change the cost-effectiveness of a specific intervention, and second, how a change in pandemic threat changes total WTP to avert pandemics with dynamic consumption and population growth process.

[17] For example, Méjean et al. apply the IAM RESPONSE model developed in Ambrosi et al. (2003).

[18] These limitations are considered more carefully in Section 6.2.

Chapter 3

Willingness To Pay to Mitigate Existential Threats

In this chapter, I develop a general model for a social planner's decision to spend current resources to mitigate existential risks (I refer to this as a 'safety investment'). I consider several different iterations of the problem to understand how various factors can change WTP for risk mitigation. I consider the safety investment as a tax levied on the annual output of current generations.[1]

To evaluate changes in aggregate utility under different levels of existential risk, I expand on the endogenous discount model applied in Méjean et al. (2020). Their model employs a social welfare function, W, and assumes that the discount on future utility is represented by the hazard rate of existential risk which determines the survival probability of a generation at t, conditional on the survival of the generation at $t-1$, $\theta_t(.)$, which is a function of the safety investment from that generation, a_t:

$$W = \sum_{t=0}^{\infty} \prod_{j=0}^{t} \theta_j(a_t) V_t(N_t, a_t, c_{ti})$$

where the value function, $V(\cdot)$, aggregates utility and is determined by the consumption, c_{ti}, of individuals, i, alive at time t, where $i = 1, 2, 3, \ldots, N_t$, and the level of safety spending at t, a_t.[2] The probability of survival of a future generation is conditional on the survival of all the generations before them, so, for example, the probability that the future generation y exists is the product of all the conditional generation survival probabilities between the present and y. We can calculate the conditional survival probability using the hazard rate of existential risk, δ:

$$\theta_t(a_t) = \frac{1}{1 + \delta_t(a_t)}$$

[1] Of course, rather than taxing current generations governments could borrow funds to invest in risk mitigation, placing the financial burden on future generations who benefit from it. However, even if this is the preferable method of funding investment risk mitigation it could still be true that a social planner is willing to sacrifice some portion of the wealth of current generations for the sake of the future – this possibility is evaluated in this thesis.

[2] The social welfare function is written here in discrete time for simplicity, though some authors prefer to define this function in continuous time.

3.1 A Numerical Experiment of Willingness to Pay for a Risk Mitigation Project

Consider a total utilitarian specification of the above social welfare function, where the value of a given generation (where each generation exists for a single period, t) is simply the sum of all individual utilities, $u_i(\cdot)$, in that generation, with the population size of a generation at t defined by N_t:[3]

$$W = \sum_{t=0}^{\infty} \prod_{j=0}^{t} \theta_j(a_t) \sum_{i=1}^{N_t} u_{ti}(c_{ti}) \tag{3.1}$$

To illustrate the WTP for a risk mitigation project, I make several simplifying assumptions on this function to illustrate the basic implications of the model. Firstly, since I am just comparing across two conditions, one in which the project is taken and one in which it is not, there are two corresponding values of θ: θ_0 under the status quo, and $\theta_1(x)$ under the project:

$$\begin{aligned} \theta_0 &= \frac{1}{1+\delta} \\ \theta_1(x) &= \frac{1}{1+(1-x)\delta} \end{aligned}$$

where δ is the status quo hazard rate, and x is the effectiveness of the project if it is pursued (the change in risk as a result of the project) which I assume is known by the social planner where $x \in (0, 1]$. Secondly, the utility function is assumed to be isoelastic:[4]

$$u_{ti}(c_{ti}) = \frac{c_{ti}^{1-\eta}}{1-\eta}$$

where η is a measure of the curvature of the utility function, and in this context represents aversion to intergenerational inequality. A social planner is said to be averse to intergenerational inequality ($\eta > 0$) if, taking two different generations that are homogeneous in intragenerational consumption but unequal in intergenerational consumption, then transferring consumption across generations such that inequality decreases is aggregate utility increasing. If we assume growth of consumption, then we can also interpret aversion to intergenerational inequality as an aversion to regressive intergenerational transfers, since any transfer from the present into the

[3] Total utilitarian because social welfare is evaluated by summing individual utility. I consider other methods of evaluating social welfare in Section 3.2.

[4] If $\eta = 1$ then the utility function is defined in log terms:

$$u(c_t) = log(c_t)$$

However, as is illustrated in Table 3.1, a value of $\eta = 1$ is rarely applied in the literature. Therefore, I do not consider it here, and only apply the above form of the utility function.

future will, *ceteris paribus*, be welfare decreasing.[5]

An issue with the application of the above utility function is that it produces negative utilities for all $\eta > 1$. A consequence of this negativity is that (3.1) yields the absurd result that increasing existential risk would increase social welfare. For example, if the utility of a potentially infinitely lived agent is -1 and constant for a discounted stream of infinite future periods, then a decrease in the discount factor (implying an increase in existential risk) from .97 to .95 would *increase* the net present value (NPV) of utility from -33.33 to -20. However, if utility were positive the results would be the reverse: with utility of $+1$, the same change in discount factor would *decrease* utility from $+33.33$ to $+20$. This problem is addressed by considering changes in utility, which is the technique applied to derive the results in this section.

Next, I assume that the population is homogeneous in intragenerational consumption, and grows at a constant rate, n, where $n \in \mathbb{R}_{\geq 0}$; therefore $\sum_{i=1}^{N} u(c_{ti})$ becomes $N_0(1+n)^t \times u(c_t)$. The assumption of homogeneous consumption is made here since the focus of this analysis is on long-term intergenerational allocation which would only be affected by heterogeneity of consumption if inequality were to grow into the future. Understanding the relationship between existential risk and future inequality may be an interesting avenue for future research, though it is beyond the scope of this thesis.

Finally, I assume consumption grows at an exogenous rate, g, and therefore $c_t = c_0(1+g)^t$. This assumption is required since the model developed only considers consumption and safety investment, and not capital allocation to growth-promoting investments as in classical growth models. Therefore, this model ignores the possibility that increasing safety spending will not only decrease consumption but also decrease growth-promoting investments. Following similar work from Méjean et al. (2020) and Martin and Pindyck (2021), I keep growth as exogenous despite these concerns since (as we will see) WTP for risk mitigation activities will realistically be at most a few percent of global income. I recognise that in the long-term a few percent of global income diverted to growth-promoting spending (rather than safety spending) could have a non-negligible effect on the global rate of economic growth; however, evaluating the mechanics of this trade-off goes beyond the scope of this thesis.[6]

In the present analysis, the social planner must decide whether to reduce present consumption (reducing utility in the present) to invest to decrease the likelihood of extinction (increasing the discount factor, θ). This spending is not only valuable for

[5] See Fleurbaey et al. (2019) and Greeves (2017).

[6] Shortly I consider the results from Aschenbrenner (2020), which assesses optimal investment in existential risk mitigation while trading off between safety spending and economic growth.

the present generation, but also for all future generations since the probability of existence of future generations is dependent on the probability that each generation before them does not go extinct.

Under these assumptions, calculating the WTP is a matter of finding the loss of consumption required such that status quo aggregate utility, W_0, is equal to the aggregate utility under risk reduction, W_1:

$$\begin{aligned} W_0 &= \sum_{t=0}^{\infty} \theta_0^t \times N_0(1+n)^t \times \frac{(c_0(1+g)^t)^{1-\eta}}{1-\eta} \\ W_1 &= \sum_{t=0}^{\infty} [\theta_1(x)]^t \times N_0(1+n)^t \times \frac{((1-a)c_0(1+g)^t)^{1-\eta}}{1-\eta} \end{aligned}$$

where a represents some proportion of income which can be used to invest in a risk mitigation project which could have otherwise been used to purchase consumption goods. The exact values that are taken by W_0 and W_1 (independent of each other) are meaningless – the values of social welfare functions only imply a ranking of social outcomes. For example, if $W_0 > W_1$ then the only information conveyed is that the status quo world is preferable to the world in which the project is invested in. Though these two functions are meaningful when considered together. For example, if $W_0 = W_1$ then this conveys information about the relationship between changes in existential risk and changes in present utility. Allowing $W_0 = W_1$, and solving for a yields the maximum WTP as a proportion of current global income for the mitigation project:[7]

$$a = 1 - \left[2 - \frac{1-\theta_1(x)(1+n)(1+g)^{1-\eta}}{1-\theta_0(1+n)(1+g)^{1-\eta}}\right]^{\frac{1}{1-\eta}}$$

To simplify this expression, I define two effective social discount rates (SDRs) which capture the discount rate on future consumption,[8] where:

$$\begin{aligned} 1+\rho_0 &\equiv \frac{(1+\delta)(1+g)^{\eta-1}}{(1+n)} \\ 1+\rho_1(x) &\equiv \frac{(1+(1-x)\delta)(1+g)^{\eta-1}}{(1+n)} \end{aligned}$$

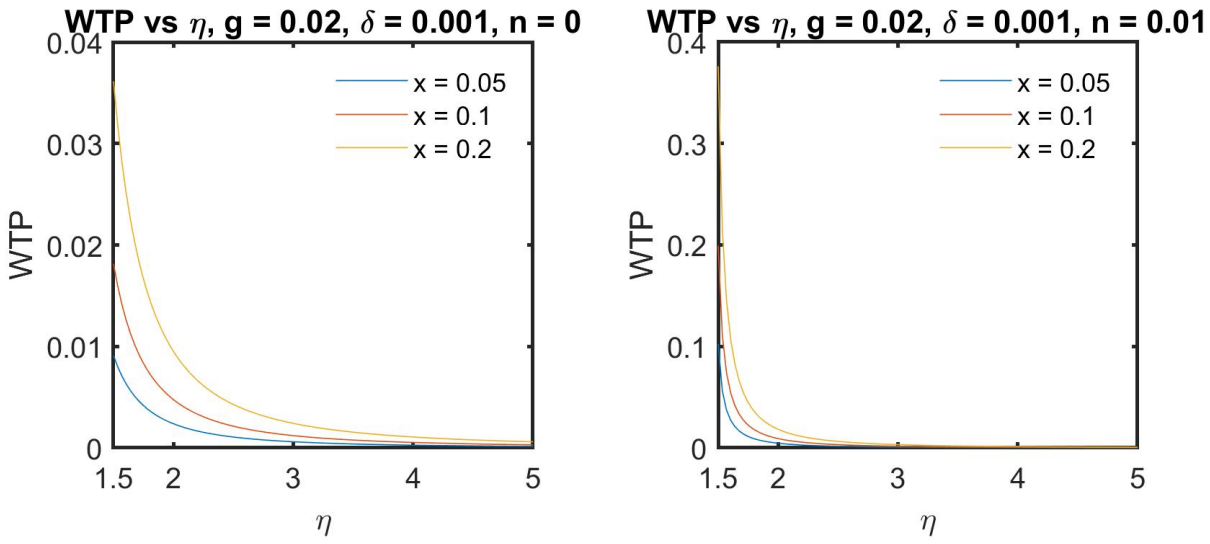
[7] See Appendix section A.2 for derivation.

[8] I defined these expressions as 'effective' SDRs since they differ slightly from the SDR derived from a discrete time social welfare function, where $(1+g)$ enters with the exponent η rather than $\eta - 1$ as above. The effective SDR is useful to simplify the expression for a, this approach is also taken in Chapter 5 though in the context of a continuous time SDR. See Appendix section A.1 for the derivation of the SDR and a review of the different forms of the SDR considered in this thesis.

and therefore:

$$a = 1 - \left[2 - \frac{1 - \left(1 + \rho_1(x)\right)^{-1}}{1 - (1 + \rho_0)^{-1}}\right]^{\frac{1}{1-\eta}}$$

Figure 3.1: Willingness to Pay for Mitigation of Existential Risk Against η



WTP for an existential threat mitigation project is graphed against aversion to intergenerational inequality, η, under three calibrations for the effectiveness of the project, x (5%, 10% and 20%), and two calibrations for the annual rate of population growth, n (0% and 1%).

As demonstrated in Figure 3.1, the most significant features of this equation for WTP are the effectiveness of the project, x, the aversion to intergenerational inequality, η, and population growth, n. In this figure I allow $g = 0.02$ (following typical calibrations in the literature), and I allow $\delta = 0.001$.[9] The importance of η in evaluating WTP for mitigation of existential risk is unfortunate since the calibrations offered in the literature for η are highly varied; see Table 3.1.

The array of values applied to η is at least partially a result of the fact that it can be interpreted, like the discount on future utility, in both normative and positive terms. For example, Weitzman (2007) and Nordhaus (2008) take a positive approach, both deriving their value for η from market activity and hence presuming that individual preferences inform the allocation decisions made by the social planner. Alternatively, Stern (2007) takes a normative approach, arguing

[9] I assume $\delta = 0.001$ following Stern (2007). I compare this calibrations against different calibrations for this parameter in Appendix C and show that the exact value of δ is not as important as other variables such as η, n or x.

Table 3.1: Calibrations for η in the Literature

Source	η
Stern (2007)	1
Cline (1993)	1.5
Groom and Maddison Pr. (2019)	1.5
Newell, Pizer, and Prest (2021)	1-2
Garnaut (2008)	1-2
Weitzman (2007)	2
Nordhaus (2008)	2
Arrow (2007)	2-3
Dasgupta (2007)	2-4
Martin and Pindyck (2021)	2-5

This table presents calibrations for η in notable works, largely from the environmental economics literature. Most entries in the table are sourced from Harrison (2010).

that the value for η ought to come from reflection upon social values towards future consumption. I do not intend to take a position on the appropriate value of η here, since I only aim to show what conditions would be necessary for existential considerations to seriously factor into cost-benefit decision making. Therefore, I consider all of the values for η which are most commonly found in the literature.

The results presented in Figure 3.1 demonstrate two particularly striking findings. Firstly, at low calibrations of η, the social planner is willing to spend a non-negligible proportion of total income (that would have otherwise been used for consumption) on risk mitigation though this value is significantly diminished at higher calibrations of η. For example, at $\eta = 1.5$ the social planner is willing to spend 3.5% and 38% of global income on mitigation for $n = 0$ and $n = 0.01$ respectively.

Secondly, relative to zero population growth, even just one percent annual population growth has a very dramatic effect on WTP while η is low, with a diminishing effect as η increases. For example, at $\eta = 1.5$, allowing the population to grow annually by 1% rather than 0% increases WTP for each value of x approximately ten-fold; however, when $\eta = 3$ the same change in population growth increases WTP by only 2% (and this decreases further as η increases). This strong effect of population growth is the result of the SDR being very close to zero at low values of η, implying that at low values of η the discounted sum of consumption for future generations is *very* large. As outlined in Chichilnisky et al. (2020), applying population growth to a social welfare function with extinction discounting can be problematic since with very low values for δ (which is the case for

extinction discounting) there is the potential for unbounded aggregate utility which occurs when the population growth rate exceeds the discount on future consumption yielding a negative SDR.

Since population growth sustained at 1% for the entire human history is exceedingly unlikely and population growth dominates the results in Figure 3.1,[10] I ignore the possibility of population growth for the rest of the following analysis until Chapter 5 where it is considered more carefully.

A further result from Figure 3.1 is that WTP decreases approximately linearly as the level of mitigation, x, decreases (at any η, for either n). For example, at $\eta = 1.5$ and and $n = 0$, the social planner is willing to spend roughly 1.8% of GDP to mitigate 10% of existential risk (roughly half the WTP to mitigate 20% of the risk), and 0.9% to mitigate 5%.

Relating the recommendations of this model to existing spending on biosecurity, according to Ord's estimates, $x = 0.2$ is equivalent to eradicating all existential threat from engineered pandemics – one-fifth of total existential risk. At $\eta = 1.5$ (assuming no population growth) a social planner is willing to spend almost $2.9 trillion USD annually (3.6% of global GDP) to totally mitigate this amount of risk, or $800 billion at $\eta = 2$ (1% of global GDP). For comparison, in 2019 the U.S budgeted $13.6 billion for biosecurity, pandemic preparedness, and 'multi-threat' health preparedness infrastructure (Watson et al., 2018). If we conservatively assume that all countries spent an equivalent per capita figure on biosecurity then this would bring global biosecurity spending to $327 billion.[11] According to this model we should be willing to spend many times more on biosecurity than we currently do under lower calibrations of η, even when ignoring the other consequences of pandemics such as lost lives and consumption which are likely to be significant (more on this in Chapter 5). However, under higher calibrations of η this conclusion differs. For example, at $\eta = 5$, to mitigate 20% of existential risk, WTP is only $48 billion (0.06% of global GDP).

3.2 Population Ethics Considerations

So far I have assumed that the marginal contribution to aggregate utility for each new individual is constant, in other words, aggregate utility is simply the sum of individual utility. One reason this approach may be undesirable is because it can lead to the 'Repugnant Conclusion' – there can be some vast number of terrible lives

[10] Population growth is expected to reach zero by the end of the century; see United Nations Population Division (2019) and Bricker and Ibbitson (2019).

[11] I assume this is a conservative figure since from 2014 only 33% of countries reported to have their health systems to minimum international standards under the International Health Regulations which are the low hanging fruit for pandemic focused biosecurity spending (Katz and Dowell, 2015).

(still worth living) which return a higher social welfare than a smaller number of very satisfying lives (Parfit, 1984). If we think that the latter outcome is preferable, or that avoiding the Repugnant Conclusion should at least carry some weight, we need to adjust the above models from their total utilitarian specification.[12] While there are a variety of suggestions for how to avoid the Repugnant Conclusion, the simplest method is to take an 'average utilitarian' view, which takes aggregate utility as the average of individual utility. Average utilitarianism has been applied both across generations, with aggregate utility taken as the average of individual utility across all of human history, and within generations, with aggregate utility taken as the sum of generation utility which is the average of individual utility within that generation. In the context of existential risk, the form of average utilitarianism applied is very important since under 'across generation' average utilitarianism future generations only increase social welfare if their welfare is higher than the existing average individual welfare. By contrast, under 'within generations' average utilitarianism, future generations continue to contribute to welfare for each new generation that exists. Either view implies that an extra individual living at average welfare (within or across generations) could contribute zero to social welfare. One might reject average utilitarianism on these grounds.

Given concerns with both these total and average utilitarian views, Ng (1986) proposes a number-dampened social welfare function, whereby social welfare lies somewhere between the average and total utilitarian evaluations (where the average is taken across all individuals who ever live). Méjean et al. (2020) develop a version of this function in the context of risky social prospects like extinction risk:

$$W = \sum_{H=0}^{\infty} P_H \big(N_H^{\beta} \sum_{t=0}^{H} \frac{N_t}{N_H} u(c_t) \big)$$

where P_H is the probability that generation H comes into existence,[13] N_H is the total number of individuals who come into existence, N_t is the population alive at time t; and β is the population ethics coefficient, where $\beta \in [0, 1]$. At $\beta = 1$, the number dampened social welfare function is equal to the total utilitarian social welfare function; at $\beta = 0$ the number dampened social welfare function is equal to the average utilitarian social welfare function. When $0 < \beta < 1$ the number dampened social welfare function evaluation is between the average and total evaluations.

[12] While avoiding the Repugnant Conclusion might seem desirable, a recent article by 29 of the leading thinkers on population ethics agree that 'avoiding the Repugnant Conclusion is not a necessary condition for a minimally adequate candidate axiology, social ordering, or approach to population ethics' (Zuber et al., 2021, p. 2). Therefore, I encourage the reader not to reject total utilitarianism outright on the grounds of the Repugnant Conclusion.

[13] P_H corresponds to θ_t in the models above.

To understand the effect of population ethics considerations on WTP to mitigate existential risk, I apply a number dampened social welfare function to the same investment decision explored in Sections 3.1. Now, we have:

$$\begin{aligned} W_0 &= \sum_{t=0}^{\infty} \theta_0^t \times N_H^{\beta-1} \times N_t \times \frac{(c_0(1+g)^t)^{1-\eta}}{1-\eta} \\ W_1 &= \sum_{t=0}^{\infty} [\theta_1(x)]^t \times N_H^{\beta-1} \times N_t \times \frac{((1-a)c_0(1+g)^t)^{1-\eta}}{1-\eta} \end{aligned}$$

Here, since population size N_t is assumed to be constant across generations, expected N_H is a function of survival probability and the population size N:

$$\mathbb{E}[N_H] = N \sum_{t=0}^{\infty} \theta^t = \frac{N}{1-\theta}$$

Assuming that the social planner takes the N_H in expectation (i.e, $N_H = \mathbb{E}[N_H]$) then allowing $W_0 = W_1$ to solve for a yields:[14]

$$a = 1 - \left[2 - \frac{(1 - \big(1+\rho_1(x)\big)^{-1})(1-\theta_1)^{\beta-1}}{(1-(1+\rho_0)^{-1})(1-\theta_0)^{\beta-1}}\right]^{\frac{1}{1-\eta}}$$

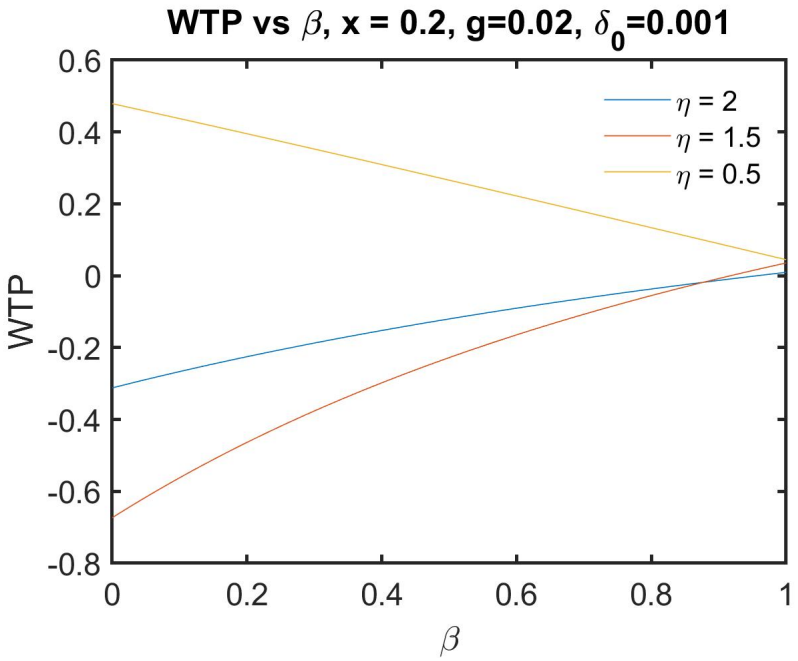
From this solution, we can see that when $\beta = 1$, a is the same as under the initial total utilitarian formulation. When $\beta \in [0, 1)$ the numerator and denominator are adjusted by a factor $(1-\theta_{0/1})^{\beta-1}$ so that the marginal social welfare change from additional future lives is decreasing. Before turning to the numerical results from this expression it is worth noting average utilitarianism significantly reduces much of the value of the future, since under this view, future generations only add to overall social welfare by elevating average individual utility if there is increased consumption in the future. This led Ord (2020) to argue that we should be motivated to reduce existential risk even if we favour an average utilitarian view, since we do expect that future generations will be richer than we are currently. However, the effects of economic growth on WTP for mitigation of existential risk are complicated; while it does increase the consumption of the future (making the expected loss of an existential catastrophe much higher), it also means that we are less willing to spend our present resources to improve future (expected) welfare if we are averse to regressive intergenerational transfers (i.e., if $\eta > 0$ while $g > 0$). The following results evaluate the validity of Ord's argument by assessing how WTP for an

[14] As in section 3.1, ρ is an effective SDR, however, here is does not include population growth:

$$\begin{aligned} 1+\rho_0 &\equiv (1+\delta)(1+g)^{\eta-1} \\ 1+\rho_1(x) &\equiv (1+(1-x)\delta)(1+g)^{\eta-1} \end{aligned}$$

existential risk mitigation project (which mitigates 20% of the total existential risk) changes with the introduction of within generations average utilitarianism, assuming constant, annual economic growth of 2% and varying aversion to intergenerational inequality. I consider two typical calibrations of this aversion, $\eta \in \{1.5, 2\}$, and a third, non-typical calibration of $\eta = 0.5$ to illustrate necessary conditions for Ord's argument to hold. Figure 3.2 presents the results from this calibration.

Figure 3.2: Willingness to Pay for Mitigation of Existential Risk Against β



WTP for an existential threat mitigation project is graphed against the population ethics co-efficient, β. At $\beta = 1$ the social welfare function is a total utilitarian specification, and at $\beta = 0$ the social welfare function is an average utilitarian specification. This relationship is presented assuming a 2% annual growth rate, a 0.1% status quo annual chance of existential collapse, and the project mitigates 20% of the existing existential risk. The figure considers three different calibrations for η.

Figure 3.2 indicates that under even very mild average utilitarian positions and $\eta \in \{1.5, 2\}$ the social planner would have to be *paid* to spend in the present to mitigate existential risk. When the social planner is significantly less averse to inequality (at $\eta = 0.5$) we see a significant increase in WTP under stronger average utilitarian positions. Therefore, we can see that Ord's claim that average utilitarianism can support a willingness to invest in mitigating existential risk only holds under an inappropriate value of aversion to intergenerational inequality. These results align with some of the general findings from Méjean et al. (2020) where it is

demonstrated for example, that average utilitarian positions tend to reduce the value of mitigating existential risk in the context of climate change. Their results suggest that average utilitarian views tend to reduce optimal climate change mitigation efforts, increasing the optimal level of global warming.

Given that the relevant economics literature tends to assume that any plausible value of η is at least 1, the above model implies that only extremely total utilitarian positions would support investment to mitigate existential risk. In the remaining analysis I assume total utilitarianism with the recognition that deviation from this position would mean that there is, at best, very little WTP to mitigate existential risk.

3.3 Optimal Spending and Economic Growth

The WTP approach taken in section 3.1 and section 3.2 identifies a simple relationship between spending to reduce existential risk and aggregated utility for a discrete spending level (to invest or not). This approach is sufficient to understand how the inclusion of existential risk as a possible consequence of pandemics can influence spending decisions for a specific project (as will be explored in Chapters 4 and 5), however there are also limitations to this approach.

The first limitation, which has already been hinted at, is that WTP analysis will not give any indication about the optimal level of spending on risk mitigation. I have suggested that deriving optimal conditions for spending on risk mitigation are particularly difficult for many existential threats since understanding these conditions requires knowledge of the relationship between safety spending and the risk level – a relationship which is unobservable.

The second is that the investment decision considered has not been modelled as an intertemporal one, where the social planner is able to maximise aggregate utility by allocating different levels of income to risk mitigation across time. The intuition behind this possibility is that in the short run a social planner may allocate resources to economic growth, even if this were to increase existential risk, so that the planner can afford to allocate a greater amount to safety in the long run.

In this section, I extend the model from section 3.1 to deal with the first limitation, providing insight into how optimal investment changes as the economy grows. I then compare these results for optimal investment against an existing model of optimal intertemporal investment in existential risk mitigation from Aschenbrenner (2020) to understand the significance of the second limitation. Ultimately, I show that the model developed in this thesis and that developed by Aschenbrenner agree qualitatively about optimal risk mitigation spending in the long run (that it should increase as we become richer), however they diverge

about the optimal short-run policy. In light of this divergence – a consequence of the second limitation – I suggest that the application of the model developed in this thesis should be limited to WTP analysis and CBA, rather than optimisation applications.

3.3.1 How Does Growth Change Optimal Safety Spending?

I begin with a social welfare function that closely resembles the functions applied above:

$$W = \sum_{t=0}^{\infty}[\theta(a)]^t \times N \times \left[\frac{((1-a)c_0(1+g)^t)^{1-\eta}}{1-\eta}\right] \tag{3.2}$$

While all the variables in (3.2) can be interpreted the same as above, now θ is a continuous function that maps the level of investment in existential risk mitigation, $a \in [0,1]$, into a unique discount factor in the range $[0,1]$. Though, the exact function, $\theta(a)$, is unknown. Here the only choice variable is a, therefore allowing $\partial W/\partial a = 0$ would allow one to solve for the value of a which maximises aggregate utility. Since the function $\theta(a)$ is unknown, maximising for a could only be done in abstract terms and is unlikely to help define an exact optimal safety investment. Instead, I apply this function to test how optimal investment should change with economic growth, and use Gronwall's inequality theorem to minimise the use of abstract functional forms in the resulting expression related to optimal investment, allowing a simpler interpretation of the results.[15]

Existing research has shown that individual health care spending has increased as a proportion of GDP across OECD countries since 1960 (Jones, 2002). One widely cited explanation for this observation is that as incomes rise the marginal utility of health care spending (to increase individual life-expectancy) increases relative to the marginal utility of consumption; as people become richer, purchasing additional years becomes more valuable relative to extra consumption (Hall and Jones, 2007). Given the similarities between allocating resources to extending individual longevity and extending species longevity, one might expect to observe a similar relationship in the context of spending to mitigate existential risk.

In the context of spending on existential risk mitigation, this question is: under what conditions do the marginal utility gains from safety spending increase relative to the marginal utility gains from further consumption while consumption is increasing? Mathematically, this corresponds to whether $\partial\left(\frac{\partial W/\partial a}{\partial W/\partial c_0}\right)/\partial c_0 > 0$.

This yields the following result:

Proposition 3.3.1. *Under (3.2), optimal a is strictly increasing in c_0 if $\theta(a) >$*

[15] The mathematical components of this process are left to Appendix section A.3.

$1 - \frac{\delta_0}{1+\delta_0}(1-a)^{\eta-1}$.

Proof. See Appendix section A.3. □

Without making any assumptions on the functional form of $\theta(a)$, this restriction is a moderate one; most plausible calibrations for δ_0, a, and $\theta(a)$ satisfy this restriction so that spending will increase with economic growth. The calibrations which conform to this restriction are illustrated in Appendix section A.4. This restriction can be simplified significantly by placing weak assumptions on the functional form of $\theta(a)$. Namely, I assume that $\theta(a)$ is a decreasing and convex function,[16] and that a reasonable candidate for the functional form is an exponential one. While there are other plausible functional forms, as we will see, an exponential form simplifies the restriction significantly while capturing the necessary features (decreasing and convex) of $\delta(a)$. Taking this form, we have $\delta(a) = \delta_0 e^{-\gamma a}$ with $\gamma \in \mathbb{R}_+$ capturing the effectiveness of safety spending at reducing risk (implying $\theta(a) = \frac{1}{1+\delta_0 e^{-\gamma a}}$).[17] Substituting this form into the restriction in Proposition 3.3.1 results in $\theta(a) > \frac{\eta-1}{\gamma(1-a)}$.[18] Furthermore, since in any optimal solution both $1-a$ and $\theta(a)$ are going to be very close to 1,[19] the restriction becomes approximately $\gamma > \eta - 1$.

It is impossible to calibrate γ using empirical data, though we can compare different calibrations of γ against intuition. For example, to mitigate 50% of the total existential risk, a social planner would have to spend roughly 75% of global GDP under $\gamma = 1$, 35% under $\gamma = 2$, and 25% under $\gamma = 3$. This relationship is presented in figure Figure 3.3.

3.3.2 Risk Mitigation as an Intertemporal Decision

Aschenbrenner (2020) demonstrates that when risk mitigation is considered as

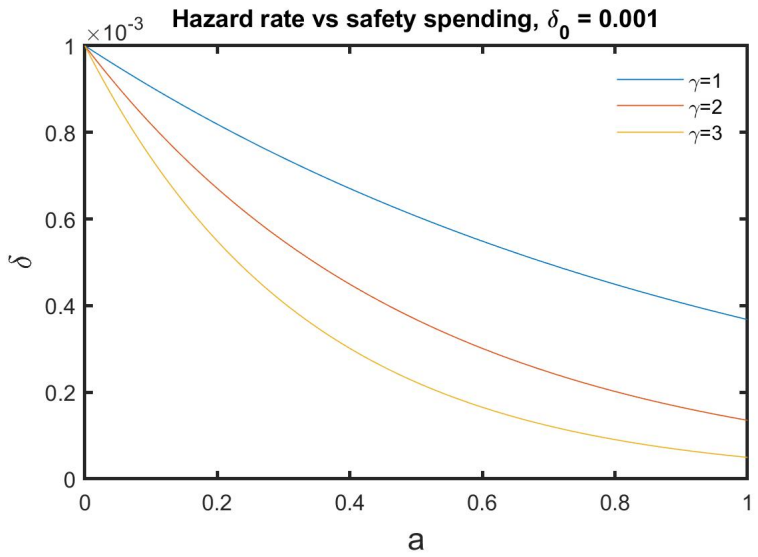
[16] This seems very likely given the probable existence of low hanging fruit for safety spending; see Leigh (2021).

[17] One possible problem with this form is that it suggests that risk diminishes with the *proportion* of income spending, rather than the *magnitude* of spending. *Prima facie* one would expect the magnitude of spending would be a more reasonable determinate of existential risk since the interventions one might think of such as 'always-on' pandemic detection systems or asteroid monitoring and deflection systems are public goods that require a constant magnitude of funding to protect a growing population or economy. However, by including proportion of income, this relationship is sensitive (even if crudely so) to the general assumption that all else the same, total existential risk will be greater in the future most notably due to the development of more dangerous technologies since funding will have to increase in absolute terms which the economy grows to maintain a constant proportion of income dedicated to safety spending implying a constant level of risk.

[18] See Appendix section A.5.

[19] So far I have assumed $\delta_0 = 0.001$ meaning that the maximum difference between $\theta(a)$ and 1 (at $a = 0$) is ≈ 0.001. Furthermore, in Section 3.1 I demonstrated that WTP to mitigate some existential risk was at most a few percent. Given that optimal spending is necessarily less than WTP, $1-a$ will be within a few hundredths of 1. Importantly, the difference between these values and 1 is very small relative to η and γ, meaning that the restriction is dominated by values of η and γ.

Figure 3.3: Calibrations of γ in $\delta(a)$



This figure depicts different calibrations of γ for $\delta(a)$, where higher values of γ imply a larger decrease in the hazard rate after an investment of a proportion of income.

an intertemporal allocation problem, the optimal investment path is to initially grow the economy (even if this increases existential risk in the short-run) and then dramatically increase spending on risk mitigation later, and consequently decrease risk in the long-run. Since the simple model developed in subsection 3.3.1 captures the decision for optimal investment for a single period by trading off present consumption for future utility (by decreasing existential risk), it is unable to demonstrate the same dynamic optimisation results as in Aschenbrenner (2020). The recommendations of this analysis and the one offered in Aschenbrenner (2020) have a similar recommendation about the long-run approach to existential risk mitigation, though the path to achieving this is notably different.

In light of the limitations of this approach, it seems that the model developed in this thesis is inadequate to inform decision making about the optimal risk mitigation policy, as attempted in subsection 3.3.1. Though, this result is useful to guide the application of the model developed in this thesis. Specifically, this result advises that this model should only be applied in contexts where the possible outcomes do not require intertemporal optimisation. I have already presented an example of such an application in section 3.1, which determined the WTP for a specific risk mitigation project, a result that does not rest on intertemporal considerations of risk. Therefore, in the subsequent chapters I only apply the model to understand

how WTP changes for an existing project after introducing existential risk, rather than considering optimal investments across generations.

3.4 The Application of the Endogenous Model to Real Existential Threats

So far I have only considered pure existential threats – a threat which, upon materialising, will always result in an existential collapse and will otherwise have no effect. In reality, any existential collapse is exceedingly likely to be the most damaging manifestation of a catastrophic event that has a distribution of impacts, including relatively mild occurrences; pandemics are certainly included in this category. Therefore, any realistic model of WTP to avoid specific catastrophic events should consider this distribution by accounting for outcomes such as lost consumption and deaths as well as the probability of existential collapse.

One plausible method (though incorrect, as I will show) to evaluate total WTP for some project to mitigate some existential threat is to simply sum the WTP for a pure existential threat (derived in section 3.1) with the WTP to mitigate the potential damages from that threat:

$$\begin{aligned} a^* &= a_{X-risk} + a_{damage} \\ &= \left[1 - \left(2 - \frac{1 - \left(1+\rho_1(x)\right)^{-1}}{1-(1+\rho_0)^{-1}}\right)^{\frac{1}{1-\eta}}\right] + \alpha\mathbb{E}[d] \end{aligned} \tag{3.3}$$

where a_{X-risk} is the WTP for existential risk mitigation calculated in section 3.1, a_{damage} is the WTP to prevent all other damages, $\mathbb{E}[d] \in (0,1)$ is the expected annual damages from some catastrophic event as a portion of global GDP,[20] and α is the proportion of the expected damages that the project will mitigate. Here, $a_{damage} = \alpha\mathbb{E}[d]$.[21]

To illustrate the bias in a^*, I consider a simplified model of an event that poses an existential threat and causes damages. As above, I consider status quo aggregate utility, W_0, and aggregate utility under some degree of threat mitigation, W_1:

$$\begin{aligned} W_0 &= \sum_{t=0}^{\infty} \theta_0^t \times N \times \frac{((1-\mathbb{E}[d])c_0(1+g)^t)^{1-\eta}}{1-\eta} \\ W_1 &= \sum_{t=0}^{\infty} [\theta_1(x)]^t \times N \times \frac{((1-(1-\alpha)\mathbb{E}[d]-a)c_0(1+g)^t)^{1-\eta}}{1-\eta} \end{aligned}$$

[20] This may include damages resulting from loss of life if a dollar value is assigned to this loss.

[21] For a benevolent social planner a_{damage} cannot exceed $\alpha\mathbb{E}[d]$, since this would make the mitigation effort more costly that the threat itself. Furthermore, $\alpha\mathbb{E}[d]$ cannot exceed a_{damage} since then the social planner could spend more than a_{damage} to mitigate α proportion of the damages and still be better off. Therefore, $a_{damage} = \alpha\mathbb{E}[d]$.

where the interpretation of each variable has not changed from above.

Allowing $W_0 = W_1$ and solving for WTP yields:

$$a = \left[1 - \left(2 - \frac{1 - \big(1 + \rho_1(x)\big)^{-1}}{1 - (1 + \rho_0)^{-1}}\right)^{\frac{1}{1-\eta}}\right](1 - \mathbb{E}[d]) + \alpha\mathbb{E}[d] \tag{3.4}$$

Proposition 3.4.1. *If W_0 and W_1 are defined, then $a^* > a$.*

Proof. See Appendix section A.6 □

The bias in a^* comes from the fact that if there is some constant loss of consumption due to (expected) damages from a catastrophe, then the WTP for existential risk mitigation should be evaluated by considering the of resources there are to consume after the catastrophe – in this case, the initial consumption multiplied by $(1 - \mathbb{E}[d])$. This difference can be observed by comparing (3.3) and (3.4).

This simple example shows that to evaluate WTP to mitigate real existential threats, it is inappropriate to simply calibrate the results from section 3.1 to an existential threat of choice and add that WTP to existing WTP estimates to mitigate the (non-existential) damages from that catastrophe. In this example, the unbiased solution for WTP could be calculated with a simple adjustment to a^*, though this simple adjustment will not always be suitable for deriving overall WTP. For example, if we include a change in population size resulting from a catastrophe then the problem becomes more complicated (as demonstrated in Chapter 5). Therefore, this result should not be interpreted as a method of generating unbiased WTP evaluations under the possibility of catastrophic damages and existential risk. Instead, Proposition 3.4.1 suggests that to evaluate WTP to mitigate real existential threats, the analyst must unite the endogenous discount model with their initial method of evaluating WTP to mitigate the non-existential outcomes (such as lost consumption or lives), rather than adding on WTP for existential risk mitigation *ex post*. This thesis gives two examples of this process by evaluating WTP for pandemic mitigation projects in Chapters 4 and 5.

3.5 Discussion

In this chapter I have demonstrated five main results from the approach used to understand WTP to mitigate existential threats which I paraphrase here. Firstly, WTP for risk mitigation strongly diminishes with aversion to intergenerational inequality, and strongly increases WTP with population growth. Secondly, only very strong total utilitarian positions support positive WTP to mitigate existential risk under conventional calibrations of aversion to intergenerational inequality. Thirdly, the change in WTP to mitigate risk increases approximately linearly with the

effectiveness of the risk mitigation intervention. Fourthly, this model has the same long-run optimal policy recommendation for risk mitigation as an intertemporal optimisation model, though differs in optimal short-run policy recommendation. Finally, in applying the endogenous discount model to an existential threat which also poses a distribution of non-existential damages, the total WTP will be less than the sum of the WTP to avert the existential threat and the non-existential threats independently.

Of these five results, the first three are apparent in the analysis of optimal GHG emissions paths in Méjean et al. (2020). However, by applying this model to a pure existential threat the effects of key variables (aversion to intergenerational inequality, population ethics specification in the social welfare function, and the effectiveness of the intervention on WTP for existential threat mitigation) are offered in isolation, rather than being intertwined with the WTP to mitigate the non-existential threats from climate change. Therefore, the present analysis offers generalised insights which could be applied to any existential threat.

The final two results provide context and motivation for the application of the model in the subsequent chapters. Specifically, the fourth result demonstrates the limits of the model developed at the beginning of this chapter, and that it cannot be easily applied to infer results about optimal policies. Therefore, I limit my application of the model in the following chapters to the analysis of WTP. The fifth result suggests that to determine WTP for mitigation of any realistic existential threat, one must unite an endogenous discount model with a model capturing the WTP to mitigate the other effects of a given catastrophe (such as lost consumption or lives), rather than summing these two WTP estimates *ex post*.

Chapter 4
Adapting Existing Cost-Benefit Analyses of Pandemic Interventions

In Chapter 3 I developed a framework to evaluate WTP to mitigate some proportion of a general existential threat. In this chapter I turn to the existential threat of pandemics, demonstrating how including existential collapse as a possible consequence of pandemics may significantly increase the cost-effectiveness of interventions to mitigate pandemic risk. This chapter considers how cost-effectiveness changes in a static environment for a specific pandemic mitigation intervention, while the following chapter considers the problem in a dynamic environment for a non-specific intervention.

4.1 Analysis of a Pandemic Intervention in Dobson et al. (2020)

Dobson et al. (2020) provide a CBA of a portfolio of interventions to mitigate the risk of future pandemics including a program to reduce zoonotic spillover, the installation of an early detection and control system, the reduction of deforestation and the eradication of wild meat trade in China – estimating implementation will cost between $22 and $31.2 billion USD annually.

To calculate the benefits from this set of interventions, the authors need an estimate for expected damages given a pandemic occurs, and the change in probability of a pandemic after the intervention. To estimate the damages, authors take the COVID-19 pandemic as a representative pandemic whereby they assume a COVID-19 scale pandemic occurs on average once in a century or once in every second century (the authors use both calibrations in their analysis).[1] The authors estimate the damages from the COVID-19 pandemic to be between be $8.1 and $15.8 trillion USD. Based on this, the authors assume that the status quo expected annual cost from pandemic damages is either between $81 and $158 billion for a once-in-a-century calibration, or between $40.5 and $79 billion for once-every-second-century

[1]The authors implicitly define a COVID-19 scale pandemic as one which causes at least the damages observed by the COVID-19 pandemic. Interestingly, if we expect to improve at managing pandemics by virtue of experience, then this would imply that for a future pandemic to be COVID-19 scale may require a greater infection fatality ratio or reproduction number.

calibration. The authors also include ancillary benefits of the intervention from reduced carbon emissions (a result of decreased deforestation) which they calculate to be $4.3 billion.

Rather than attempting to predict the change in pandemic probability resulting from their intervention without any empirical evidence,[2] the authors instead opt to work backwards from the estimated costs and benefits, and the calculate the 'break-even' probability – the required change in probability of future pandemics resulting from the intervention such that the costs would be exactly equal to the expected benefits. One can calculate the break-even probability by solving the following equation (which equates status quo expected costs with expected cost under the intervention) for P_1:

$$P_0 \times D = P_1 \times D - C$$

where P_1 is the post intervention break-even probability, C is the cost of the intervention (net of ancillary benefits), D is the damages given a COVID-19 scale pandemic,[3] and P_0 is the initial probability of a COVID-19 scale pandemic. The

Table 4.1: Dobson et al. (2020) Results

P_0	C(M$)	D(M$)	P_1	$\%\Delta P$
0.01	31,211	11,506,430	0.0073	27.12
0.01	26,904	11,506,430	0.0077	23.38
0.01	17,686	15,790,780	0.0089	11.20
0.01	31,211	15,790,780	0.0080	19.77
0.01	17,686	8,126,938	0.0078	21.76
0.005	17,686	15,790,780	0.0039	22.40
0.01	31,211	8,126,938	0.0062	38.40
0.005	31,211	15,790,780	0.0030	39.53
0.005	17,686	8,126,938	0.0028	43.52[4]
0.005	31,211	8,126,938	0.0012	76.81

This table presents the results from the Dobson et al. (2020) analysis of the required effectiveness of a pandemic intervention under various calibrations of pandemic risk, costs and damages. Where P_0 is the initial pandemic probability, $C(M\$)$ is the cost of the intervention in millions of US dollars, $D(M\$)$ is the damages of a COVID-19 scale pandemic in millions of US dollars, P_1 is the required probability of a pandemic after the intervention to break even, and $\%\Delta P$ is the change in P in percentage terms.

[2] The difficulty of predicting the effects of a pandemic intervention on pandemic probability has not stopped World Bank (2012) nor Millet and Snyder-Beattie (2017) who proceed in a CBA with subjective estimates of the change in this probability.

[3] The authors include the loss of life in the damage term using the value of a statistical life estimates.

[4] Dobson et al. (2020) report this value as 45.52, though my own calculations suggest this value

authors calculate break-even probabilities for low, mid-range and high estimates of costs and damages, and low and high estimates for the status quo probability of a COVID-19 scale pandemic. Their results are presented in Table 4.1.

These results show that under different calibrations of costs and expected damages, this intervention can be as low as 11.2% effective, or as high as 76.8% effective to break even. While this analysis is focused exclusively on avoiding future COVID-19 scale pandemics, we ought to expect that if this intervention is also effective at mitigating the probability of more severe pandemics (which I assume to be the case, though more on this in the following section) then this required break-even probability would be lower, perhaps even significantly so.

4.2 Introducing Existential Risk

To understand the effect of the most extreme of pandemic outcomes – existential collapse – on the resulting break-even probability, I employ the approach developed in Chapter 3 where changes in existential risk, resulting from the proposed intervention, are captured with an endogenous discount factor. I consider three different calibrations for existential risk conditional on the occurrence of a pandemic of at least COVID-19 scale: 1 in 1,000, 1 in 10,000, and 1 in 100,000. These calibrations are not intended to be precise estimates; rather, they represent the range of orders of magnitude for which we observe significant changes in break-even probability as a result of existential considerations.

Multiplying the conditional calibrations by the status quo pandemic probability from Dobson et al. (2020) of 0.01 and 0.005 yields 6 different calibrations for annual existential risk from pandemics. These calibrations are presented in Table 4.2.

Table 4.2: Calibrating δ

P_0	$P(E\|P)$	δ_p
0.01	$0.001 \equiv P(H)$	10^{-5}
	$0.0001 \equiv P(M)$	10^{-6}
	$0.00001 \equiv P(L)$	10^{-7}
0.005	$0.001 \equiv P(H)$	5×10^{-6}
	$0.0001 \equiv P(M)$	5×10^{-7}
	$0.00001 \equiv P(L)$	5×10^{-8}

This table presents the calibrations considered for status quo existential risk from pandemics, δ_p, which is calculated by multiplying status quo pandemic probability, P_0, by the probability of extinction given a pandemic, $P(E|P)$.

So far I have not distinguished between existential risk of pandemics arising

is as reported above.

from natural compared to engineered or leaked sources. While Dobson et al. (2020) focus on the prevention of naturally occurring pandemics, elements of their proposed intervention are likely to be effective at mitigating leaked or engineered pandemic risk, such as their proposed early pandemic detection system.[5] Therefore, the true change in existential risk that can be mitigated from this intervention (the endogenous existential risk), δ_p, is the sum of changes in naturally occurring pandemic risk, $\delta_{p(N)}$ and existential pandemic risk from engineered ($\delta_{p(E)}$) or lab leaked sources ($\delta_{p(L)}$) which is also mitigated by this intervention:

$$\Delta\delta_p = \Delta\delta_{p(N)} + \Delta\delta_{p(E)} + \Delta\delta_{p(L)}$$

A value for $\Delta\delta_p$ is essential to calculating welfare under the mitigation intervention. To derive this value I assume that the proportion of pandemics that are existential threats is fixed, and only the probability of pandemics is endogenous so $\Delta\delta_{p(N)} = A \times \Delta P$, where A is some constant. In this case, $A = P(E|P)$ which is the probability of extinction, given a pandemic occurs. However, there are good reasons why this assumption may not hold. For example, perhaps the worst pandemics cannot be prevented as effectively as milder pandemics, so that as we mitigate pandemic risk the only pandemics which could still happen are the most severe. This would imply that the change in existential risk from pandemics would be less than the change in the probability of a pandemic after the mitigation intervention. It is also plausible that a government could adopt a targeted intervention to mitigate existential threats, so the change in existential risk is greater than the change in the probability of a pandemic. In the absence of strong evidence as to how the change in existential risk from pandemics relates to the annual probability of a pandemic, I continue with the assumption that $\Delta\delta_p$ and ΔP are directly proportional and recognise this as a potential limitation of the following analysis.

The same argument cannot be made for deriving values for $\Delta\delta_{p(E)}$ and $\Delta\delta_{p(L)}$, since there is no reason to expect that the probability of existential risk from engineered or lab-leaked pandemics is proportional to the frequency of natural pandemics, P. Without any empirical means of calibrating these values, I take several calibrations at different orders of magnitude for δ_p to illustrate how the break-even probability changes under different levels of risk. These orders of magnitude cover the estimate of existential threat from pandemics in Ord (2020). For example, Ord places $\delta_{p(N)}$ at 10^{-6} corresponding to the $P(M)$ calibration from Table 4.2 when $P_0 = 0.01$, though he places $\delta_{p(E)}$ 333 times higher at 3.33×10^{-4}. Therefore, under

[5] For example, the World Health Organization claims their International Health Regulations, which include an early pandemic detection system, are effective against natural occurrences of epidemics or pandemics, and accidental or deliberate release of biological or chemical agents (World Health Organization, 2005).

Ord's estimates only 2.7% of the total existential risk from engineered pandemics would have to be mitigated by the proposed intervention to arrive at the calibration for δ_p corresponding to $P(H)$ when δ_p when $P_0 = 0.01$ according to Table 4.2.[6]

To calculate the new break-even probabilities under these calibrations for extinction risk, I define two social welfare functions that aggregate utility under the status quo, W_0, and if the pandemic mitigation project is undertaken, W_1:

$$W_0 = \sum_{t=0}^{\infty} \frac{1}{(1+\delta_{p(0)}+\delta_{p\prime})^t} \times N \times \frac{\big((c_0 - D \times P_0)(1+g)\big)^{1-\eta}}{1-\eta}$$

$$W_1 = \sum_{t=0}^{\infty} \frac{1}{(1+\delta_{p(1)}+\delta_{p\prime})^t} \times N \times \frac{\big((c_0 - D \times P_1 - C)(1+g)\big)^{1-\eta}}{1-\eta}$$

where $\delta_{p(0/1)}$ is the endogenous extinction risk when the intervention is not taken ($p(0)$), or is taken ($p(1)$); and $\delta_{p\prime}$ is the exogenous extinction risk from non-pandemic events.[7]

4.3 Results and Discussion

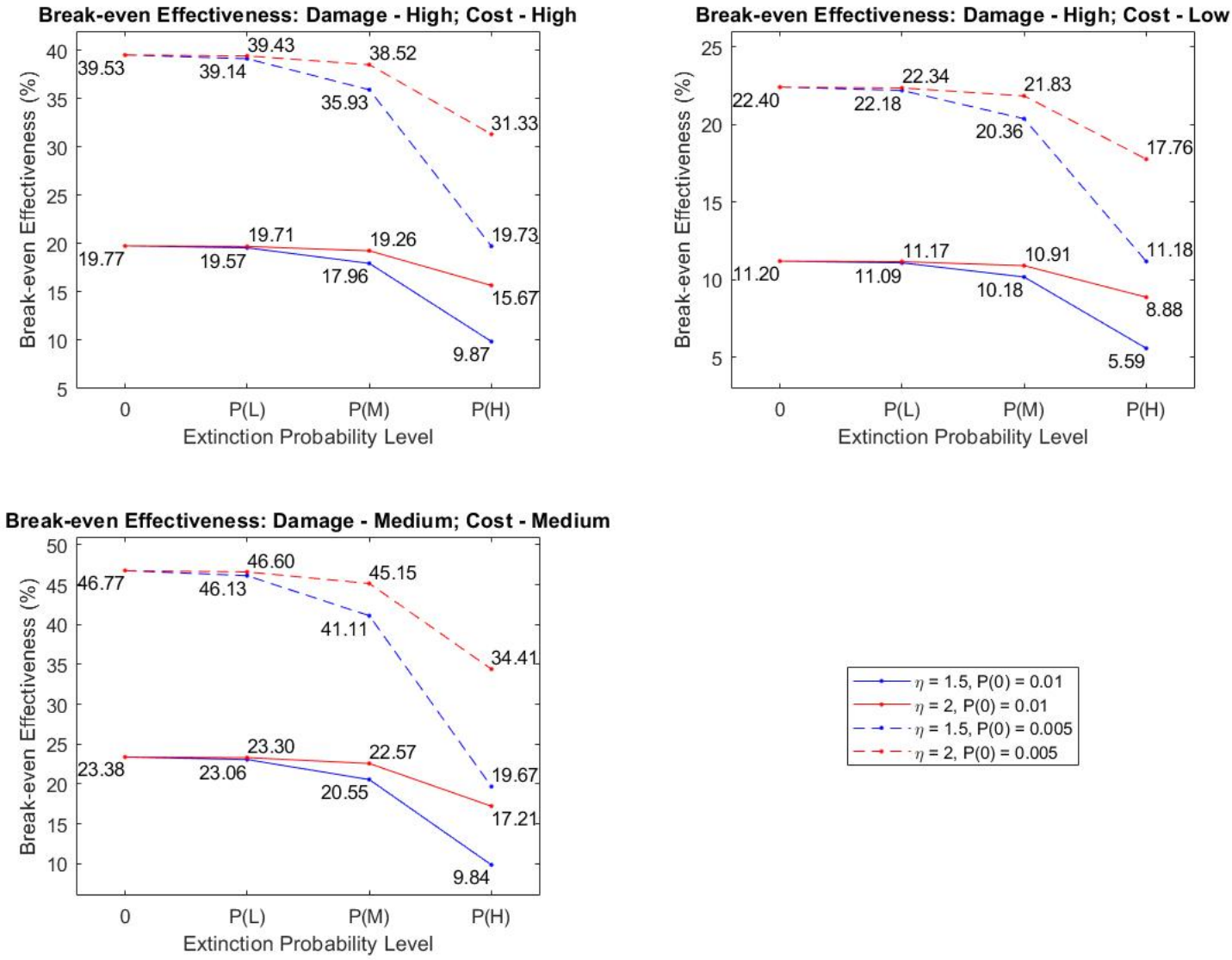
This equation is solved under both $\eta = 1.5$ and $\eta = 2$ corresponding to the typical values applied in the literature. Figure 4.1 presents selected results from this calculation to illustrate the general pattern of break-even probability under $P(L)$, $P(M)$, or $P(H)$, corresponding to values in Table 4.2. These results are presented in full in Table 4.3.

These figures illustrate three general points. Firstly, the relative patterns across values of break-even effectiveness across different damage and cost calibrations presented in the three graphs of Figure 4.1 are approximately similar; this implies that the effect of existential considerations on break-even probability is relatively constant across calibrations of costs and damages. Secondly, as in the general case presented in Chapter 3, higher values of η decrease the WTP to take actions to preserve future utility. Finally, under $P(L)$ and $P(M)$ (corresponding to a one in one thousand and a one in ten thousand chance of extinction given a pandemic occurs that of at least COVID-19 scale) existential considerations have very limited impact on the resulting break-even probabilities. Since, *ex ante*, any estimate of the effectiveness of an intervention is highly subjective, if an analyst believes that there is less than a one in one thousand chance of a pandemic causing an existential collapse, conditional on a pandemic occurring, their analysis will be entirely dominated by

[6] $\delta_p = \delta_{p(N)} + 0.027\delta_{p(E)}$ when $\delta_p = 0.01 \times P(H)$, $\delta_{p(N)} = 1/1,000,000$, and $\delta_{p(E)} = 1/3,000$.

[7] I solve for P_1 when $W_0 = W_1$ using Wolfram Mathematica, though this generally yields multiple solutions for P_1, however there is only ever a single positive solution. Negative solutions can be ignored since this implies the intervention could *increase* the probability of pandemics, which is assumed to be impossible.

Figure 4.1: Break-Even Effectiveness Under Various Calibrations of δ, η, and $P(0)$



This figure presents the change in break-even effectiveness of the pandemic intervention proposed by Dobson et al. (2020) under selected calibrations for the expected cost of the pandemic intervention, and damage of a pandemic. Full results are presented in table form in Table 4.3

non-existential considerations such as the cost of the intervention and expected damages (excluding extinction) of a pandemic. However, if the analyst believes that we face risk in the range of $P(H)$ then the analyst certainly should factor in extinction risk; *ceteris paribus*, the possibility of extinction would make the intervention roughly twice as cost-effective under $\eta = 1.5$, and roughly 25% more cost-effective under $\eta = 2$. One might be tempted to interpret these figures as implying that a high level of existential threat has a stronger effect under lower status quo pandemic probability ($P_0 = 0.005$ rather than $P_0 = 0.01$). However, this is merely an artifact of the relative impact of extinction to damages, where under $P_0 = 0.005$ extinction risk represents a comparatively greater proportion of the total damages than under $P_0 = 0.01$. Therefore I encourage the reader to avoid comparing the impact of extinction probability across $\eta = 1$ and $\eta = 2$.

As a rule of thumb consider an *ex ante* belief in the order of magnitude of $P(H)$ to be the minimum conditions for existential considerations to have a relevant impact on cost-effectiveness estimates. I have included $P(H)$ here not to represent an upper bound estimate, but instead to demonstrate the necessary conditions to see a significant shift in break-even probability after including existential risk in the model.

This is because it is very difficult to anticipate exactly what the change in pandemic probability from engineered pandemics will be if this intervention which targets naturally arising pandemics, this is particularly precarious since the break-even probability is very sensitive to the changes in engineered pandemic probability since the base rate of risk from engineered pandemics is significantly higher (as per Ord's estimates).

Consequently, if we expect that this intervention has no impact on the probability of existential threat from engineered pandemics (I have suggested this seems unlikely), then including existential risk in the model has no practically important impact on the break-even probability. However, if we expect that this intervention could reduce even a small percentage of the existential risk from engineered pandemics (shifting the calibration to the $P(H)$ level if we accept Ord's baseline risk estimates), then the break-even probability is reduced by an amount of practical importance.

Table 4.3: Affect of Existential Considerations on Break-Even Probability Compared to Dobson et al. (2020) Results

P_0	D($M)	C($M)	$\%\Delta P(0)$	$\eta = 1.5$			$\eta = 2$		
				$\% \Delta P(L)$	$\% \Delta P(M)$	$\% \Delta P(H)$	$\% \Delta P(L)$	$\%\Delta P(M)$	$\% \Delta P(H)$
0.01	15,790,783	31,211	19.77	19.57	17.96	9.87	19.71	19.26	15.67
		26,904	17.04	16.87	15.49	8.51	16.99	16.61	13.51
		17,686	11.20	11.09	10.18	5.59	11.17	10.91	8.88
	11,506,430	31,211	27.12	26.76	23.84	11.41	27.03	26.19	19.96
		26,904	23.38	23.06	20.55	9.84	23.30	22.57	17.21
		17,686	15.37	15.16	13.51	6.47	15.32	14.48	11.31
	8,126,938	31,211	38.40	37.67	32.14	13.02	38.21	36.55	25.46
		26,904	33.11	32.47	27.71	11.22	32.94	31.51	21.95
		17,686	11.20	11.09	10.18	5.59	11.17	10.91	8.88
0.005	15,790,783	31,211	39.53	39.14	35.93	19.73	39.43	38.52	31.33
		26,904	34.08	33.74	30.97	17.01	33.99	33.21	27.01
		17,686	22.40	22.18	20.36	11.18	22.34	21.83	17.76
	11,506,430	31,211	54.25	53.51	47.68	54.06	53.51	52.27	39.92
		26,904	46.77	46.13	41.11	19.67	46.60	45.15	34.41
		17,686	30.74	30.32	27.02	12.93	30.63	29.68	22.62
	8,126,938	31,211	76.81	75.34	64.27	26.03	76.42	73.09	50.92
		26,904	66.21	64.95	55.41	22.44	65.88	63.01	43.89
		17,686	43.52	42.69	36.42	14.75	43.30	41.42	28.85

This table presents the results of the impact of existential considerations on estimates of the break-even effectiveness of the pandemic intervention proposed by Dobson et al. (2020); where P_0 is the status quo pandemic probability, $D(M\$)$ is the damages of a COVID-scale pandemic in millions of US dollars, $C(M\$)$ is the cost of the proposed intervention in millions of US dollars, $\%\Delta P(0)$ is the required percentage change in pandemic probability to break even under no existential risk from pandemics, $\%\Delta P(L)$, $\%\Delta P(M)$, $\%\Delta P(H)$ are the required percentage change in pandemic probability to break even under low, medium and high calibrations of existential risk from pandemics, and aversion to intergenerational inequality, η, is calibrated to 1.5 and 2.

Chapter 5
Dynamic Model of Willingness to Pay to Avert Pandemics

The previous chapter evaluated WTP for a specific intervention that recognises lost consumption and lost lives from pandemics, where it is assumed that a pandemic introduces a static, predictable loss – a pandemic is not a shock, but a permanent and unchanging burden on consumption and lives. Furthermore, in evaluating the welfare loss from deaths due to a pandemic the previous chapter only considered the welfare loss due to deaths but did not consider the potential gains from wealth redistribution after an individual has died. These gains may partially offset the total welfare loss associated with lost life.

The present chapter relaxes these assumptions by considering a pandemic shock in a dynamic environment, and reports the WTP to avoid such a shock; and also considers the welfare contribution from redistributed wealth after a death. To evaluate this WTP, I adopt a model of WTP to avert catastrophes developed in Martin and Pindyck (2021), and adapt this model to include an endogenous discount rate using the framework developed in Chapter 3.

5.1 Dynamic Model of Willingness To Pay to Avert a General Existential Threat

Similarly to the models developed in Chapters 3 and 4, the model developed in Martin and Pindyck (2021) evaluates WTP for mitigation of a threat by comparing social welfare under the status quo to social welfare under additional safety spending. Here I outline the main features of their model, and how it can be adjusted to evaluate catastrophes under the possibility of existential collapse. I leave some of the details of this model to Appendix section A.7, and note where I have relegated such information.

Martin and Pindyck define status quo welfare, W_0, in terms of the utility from consumption, $u(C_t)$, and the number of people who live after a catastrophe, N_t. The authors assume that the population is homogeneous in age and consumption so that the welfare loss associated with lost life is the same for any member of the population. They also consider the change in welfare resulting from deaths from the

given catastrophe (the number of which is $N_t^* - N_t$, where N_t^* is the total number of people had there been no catastrophe) whereby the welfare contributions from each lost life is defined as $v(C_t)$. Importantly, there are both positive and negative contributors to welfare within $v(C_t)$. The positive factor of utility from lost lives is that wealth, w, is distributed to the living via bequests. The negative factor is, of course, the loss of life. This loss is calculated using the value of a statistical life (VSL) which is defined here as a multiple, s, of the lifetime income, so, $s \times w$ is the total VSL.[1]

Combining welfare contributions from both the living and the dead, the authors arrive at the status quo expected social welfare function:

$$W_0 = \mathbb{E}\left\{ \int_0^\infty e^{-\delta t} \left[N_t u(C_t) + (N_t^* - N_t) v(C_t) \right] dt \right\} \tag{5.1}$$

where δ is the pure time preference. Substituting an isoelastic utility function into (5.1) yields the following definition of status quo welfare:

$$W_0 = \mathbb{E}\left\{ \int_0^\infty e^{-\delta t} \left[N_t \frac{C_t^{1-\eta}}{1-\eta} + (N_t^* - N_t) \frac{(C_t \varepsilon)^{1-\eta}}{1-\eta} \right] dt \right\} \tag{5.2}$$

where, to simplify the equation, ε is the proportion of utility of the dead relative to the utility of the alive;[2] and η is interpreted as the aversion to intergenerational inequality as in the previous chapters. The authors offer no interpretation of δ, but for the purposes of considering existential risk I interpret this value as the hazard rate for existential risk. Here $\delta = \delta_p + \delta_{p'}$ where δ_p is the endogenous existential risk (existential risk from pandemics), and $\delta_{p'}$ is the exogenous extinction risk.

Another plausible method of including existential risk in this model would be to apply a function describing the likelihood and distribution of death catastrophes which includes the possibility of death catastrophes resulting in human extinction. However, given that the positive contribution to utility from the dead rests on the ability of other people to remain alive to receive bequests (not relevant to an extinction scenario) then it is preferable to separate a death catastrophe from an existential collapse by applying endogenous discounting.

To introduce dynamic consumption levels subject to change from catastrophic threats, the authors apply the following consumption evolution process:

$$C_t = e^{c_t} = e^{gt - \sum_{k=1}^{Q(t)} \phi_k}$$

[1] To derive the value for s, Martin and Pindyck take a standard approach to calculate VSL, where empirical studies are used to evaluate how much compensation individuals are willing to accept to increase their risk of death. See Appendix subsection A.7.1 for further discussion of VSL and derivation of the relationship between $v(C_t)$ and $u(C_t)$.

[2] $\varepsilon = \left(1 + s(\eta - 1)\right)^{\frac{1}{1-\eta}} < 1$. This equation is derived in Appendix subsection A.7.1.

where C_0 is scaled to 1, g is the natural rate of consumption growth, and $Q(t)$ is the Poisson counting process where upon realisation of the kth catastrophic event, consumption falls by $e^{-\phi_k}$. To introduce dynamic population levels subject to change from catastrophic threats, the authors apply the following population evolution process:

$$N_t = e^{n_t} = e^{nt - \sum_{k=1}^{X(t)} \psi_k}$$

where N_0 is scaled to 1, nt is the natural rate of population growth, and $X(t)$ is the Poisson counting process where upon realisation of the kth catastrophic event, population falls by $e^{-\psi_k}$.[3]

Finally, the authors introduce two cumulant-generating functions (CGFs), $\boldsymbol{\kappa}_N$ and $\boldsymbol{\kappa}_C$, to describe the Poisson counting processes for consumption and deaths, where

$$\mathbb{E}[N_t C_t^{1-\eta}] = \mathbb{E}N_t \mathbb{E}C_t^{1-\eta} = e^{\boldsymbol{\kappa}_N(1)t} \cdot e^{\boldsymbol{\kappa}_C(1-\eta)t}$$

and

$$\mathbb{E}[(N_t^* - N_t)\varepsilon^{1-\eta} C_t^{1-\eta}] = (e^{\boldsymbol{\kappa}_N^*(1)t} - e^{\boldsymbol{\kappa}_N(1)t})\varepsilon^{1-\eta} e^{\boldsymbol{\kappa}_C(1-\eta)t}$$

and substituting into (5.2) yields:

$$W_0 = \frac{1}{1-\eta}\left\{\frac{1-\varepsilon^{1-\eta}}{\delta - \boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)} + \frac{\varepsilon^{1-\eta}}{\delta - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C(1-\eta)}\right\} \tag{5.3}$$

To evaluate the WTP to mitigate all pandemic risk I compare status quo welfare, W_0, with the welfare under no pandemic risk, W_1. To calculate W_1, I consider three substitutions into (5.3):

(a) $\boldsymbol{\kappa}_N(1)t$ is replaced with $\boldsymbol{\kappa}_N^*(1)t$, where $\boldsymbol{\kappa}_N^*(1)t$ is the population evolution process under no pandemic threat.[4]

(b) $\boldsymbol{\kappa}_C$ is replaced with $\boldsymbol{\kappa}_C^*$, where $\boldsymbol{\kappa}_C^*$ is the consumption evolution process under no pandemic threat.

(c) δ is replaced with $\delta_{p'}$, where $\delta_{p'}$ is the exogenous hazard rate, reflecting extinction risk from all non-pandemic threats.[5]

Making the substitutions described in (a), (b), and (c) above to evaluate welfare under complete pandemic mitigation considering death, consumption loss, and

[3] For further detail on these functions, see the original paper.

[4] Martin and Pindyck (2021) evaluate WTP to mitigate a pandemic threat, however only consider welfare effects from a decrease in population by making substitution (a), whereas I consider substitutions (a), (b), and (c) to consider a wider range of consequences from pandemics.

[5] See Appendix subsection A.7.2 for a discussion of some of the issues with uniting the endogenous discount model with the Martin and Pindyck (2021) model.

existential risk yields:

$$W_1 = W_{c,d,x} = \frac{(1-a_{c,d,x})^{1-\eta}}{1-\eta}\left\{\frac{1}{\delta_{p\prime} - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C^*(1-\eta)}\right\} \tag{5.4}$$

where $a_{c,d,x}$ is the proportion by which income must be reduced such that $W_1 = W_0$. Equating (5.4) and (5.5) and solving for WTP yields:

$$a_{c,d,x} = 1 - \left[\frac{\big(\delta_{p\prime} - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C^*(1-\eta)\big)\big(\delta - (1-\varepsilon^{1-\eta})\boldsymbol{\kappa}_N^*(1) - \varepsilon^{1-\eta}\boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)\big)}{\big(\delta - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C(1-\eta)\big)\big(\delta - \boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)\big)}\right]^{\frac{1}{1-\eta}} \tag{5.5}$$

Following Martin and Pindyck (2021), an alternative specification of the model is to exclude pandemics as a threat to consumption. This approach seems inappropriate given the immense economic burden of the COVID-19 pandemic,[6] though I include this specification for comparison. Therefore, Making the substitutions described in (a) and (c) above to evaluate welfare under pandemic mitigation considering death and existential risk yields:

$$W_1 = W_{d,x} = \frac{(1-a_{d,x})^{1-\eta}}{1-\eta}\left\{\frac{1}{\delta_{p\prime} - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C(1-\eta)}\right\} \tag{5.6}$$

and equating (5.4) and (5.6) yields:

$$a_{d,x} = 1 - \left[\frac{\big(\delta_{p\prime} - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C(1-\eta)\big)\big(\delta - (1-\varepsilon^{1-\eta})\boldsymbol{\kappa}_N^*(1) - \varepsilon^{1-\eta}\boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)\big)}{\big(\delta - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C(1-\eta)\big)\big(\delta - \boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)\big)}\right]^{\frac{1}{1-\eta}} \tag{5.7}$$

To compare solutions for WTP – (5.5) and (5.7) – against a base case where existential risk is not considered, I also derive WTP to avert a population and consumption catastrophe, $a_{c,d}$, as well as a catastrophe which only reduces population, a_d. To derive $a_{c,d}$, first I make the substitutions described in (a) and (b) above, yielding:

$$W_1 = W_{c,d} = \frac{(1-a_{c,d})^{1-\eta}}{1-\eta}\left\{\frac{1}{\delta - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C^*(1-\eta)}\right\} \tag{5.8}$$

Then equating (5.4) and (5.8) , I solve for $a_{c,d}$:[7]

$$a_{c,d} = 1 - \left[\frac{\big(\delta - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C^*(1-\eta)\big)\big(\delta - (1-\varepsilon^{1-\eta})\boldsymbol{\kappa}_N^*(1) - \varepsilon^{1-\eta}\boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)\big)}{\big(\delta - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C(1-\eta)\big)\big(\delta - \boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)\big)}\right]^{\frac{1}{1-\eta}} \tag{5.9}$$

[6] The International Monetary Fund (IMF) reports that relative to projected growth estimates before the COVID 19 pandemic, global output decreased by 7% in 2020 (IMF, 2021).

[7] Both (5.9) and (5.11) are derived in Martin and Pindyck (2021).

To derive a_d, I make just substitution described in (a) above:

$$W_1 = W_d = \frac{(1-a_d)^{1-\eta}}{1-\eta}\left\{\frac{1}{\delta - \boldsymbol{\kappa}_N^*(1) - \boldsymbol{\kappa}_C(1-\eta)}\right\} \tag{5.10}$$

Then equating (5.4) and (5.10) , I solve for a_d:

$$a_d = 1 - \left[\frac{\delta - \boldsymbol{\kappa}_N(1)\varepsilon^{1-\eta} - \boldsymbol{\kappa}_N^*(1)(1-\varepsilon^{1-\eta}) - \boldsymbol{\kappa}_C(1-\eta)}{\delta - \boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)}\right]^{\frac{1}{1-\eta}} \tag{5.11}$$

To evaluate WTP for a specific event (such as a pandemic) we need to define the distribution of impacts from pandemics. Following the literature on catastrophic risk, Martin and Pindyck assume economic damage and lost lives from catastrophic events follow a power law distribution, resulting in the expected survival proportion, z_x of a representative good, x:[8]

$$\mathbb{E}[z_x] = \frac{\beta_x}{\beta_x + 1}$$

where β_x is the impact parameter for good x where higher values imply a smaller expected loss.

Given the distributions of β_d and β_c, the authors define the CGF for consumption, $\boldsymbol{\kappa}_c$, and lives, $\boldsymbol{\kappa}_N$, for the status quo consumption and population growth processes:

$$\begin{aligned} \boldsymbol{\kappa}_C(1-\eta) &= g(1-\eta) - \frac{\lambda_c(1-\eta)}{\beta_c + (1-\eta)} \\ \boldsymbol{\kappa}_N(1) &= n - \frac{\lambda_d}{\beta_d + 1} \end{aligned}$$

where λ_c represents the mean arrival frequency for a consumption catastrophe, and λ_d represents the mean arrival frequency for a death catastrophe. Since I am evaluating WTP to mitigate a specific event that simultaneously inflicts a death and consumption catastrophe, I take $\lambda = \lambda_c = \lambda_d$. To evaluate welfare under no pandemic catastrophe I let $\lambda = 0$, and therefore we have:

$$\begin{aligned} \boldsymbol{\kappa}_C^*(1-\eta) &= g(1-\eta) \\ \boldsymbol{\kappa}_N^*(1) &= n \end{aligned}$$

Furthermore, following Martin and Pindyck, I substitute in the effective social

[8] In this context x represents both lives and wealth.

discount rate (SDR). Under the status quo, I define the SDR:[9]

$$\rho \equiv \delta - n + g(\eta - 1)$$

and under conditions of no existential threat from pandemics, I define the SDR:

$$\rho_{p\prime} \equiv \delta_{p\prime} - n + g(\eta - 1)$$

Substituting ρ and $\rho_{p\prime}$ into (5.5), (5.7), (5.9), and (5.11) yields:

$$a_{c,d,x} = 1 - \left[\frac{\rho_{p\prime}(\rho - \lambda_c' + \lambda_d'\varepsilon^{1-\eta})}{(\rho - \lambda_c')(\rho - \lambda_c' + \lambda_d')}\right]^{\frac{1}{1-\eta}} \tag{5.12}$$

$$a_{d,x} = 1 - \left[\frac{(\rho_{p\prime} - \lambda_c')(\rho - \lambda_c' + \lambda_d'\varepsilon^{1-\eta})}{(\rho - \lambda_c')(\rho - \lambda_c' + \lambda_d')}\right]^{\frac{1}{1-\eta}} \tag{5.13}$$

$$a_{c,d} = 1 - \left[\frac{\rho(\rho + \lambda_d'\varepsilon^{1-\eta} - \lambda_c')}{(\rho - \lambda_c')(\rho + \lambda_d' - \lambda_c')}\right]^{\frac{1}{1-\eta}} \tag{5.14}$$

$$a_{d} = 1 - \left[\frac{\rho + \lambda_d'\varepsilon^{1-\eta} - \lambda_c'}{\rho + \lambda_d' - \lambda_c'}\right]^{\frac{1}{1-\eta}} \tag{5.15}$$

where, following Martin and Pindyck, for simplicity I define impact-adjusted mean arrival frequencies of consumption and death catastrophes:

$$\lambda_c' = \frac{\lambda(\eta - 1)}{\beta_c + 1 - \eta}$$
$$\lambda_d' = \frac{\lambda}{\beta_d + 1}$$

which recognises the welfare equivalence between increasing an impact parameter $\beta_{c/d}$ which reduces the expected impact, and reducing λ which extends the expected arrival frequency of a catastrophe.[10]

The four solutions for WTP derived in this section allow us to compute how existential considerations affect WTP for complete mitigation of pandemic threat when existential considerations are ignored; where the difference between $a_{c,d,x}$ and $a_{c,d}$ yields the impact of existential considerations assuming that pandemics pose a threat to consumption and life, and the difference between $a_{d,x}$ and a_d yields the impact of existential considerations assuming that pandemics only pose a threat to life. This impact is evaluated numerically in section 5.3, but first I calibrate the above equations for WTP to pandemics in the following section.

[9] This discount rate differs slightly from the social discount rate suggested in Chapter 3 which was derived in Ramsey (1928), where $\rho = \delta + g\eta$. This is because this model considers population growth, and includes $g(\eta - 1)$ rather than $g\eta$ to simplify the results for $a_{c,d,x}$, $a_{c,d}$, $a_{d,x}$, and a_d.

[10] While I have allowed $\lambda = \lambda_c = \lambda_d$ this does not imply $\lambda' = \lambda_c' = \lambda_d'$ since the latter include impact parameters which are not the same for consumption and death catastrophes.

5.2 Calibration to Pandemics

5.2.1 Frequency and Severity of Pandemics

As I have already highlighted, calibrating pandemic frequency to past data is problematic due to the likely failure of past-future symmetry, however doing so can provide a useful starting point for estimates of the frequency of future pandemics. Examples of pandemics since the beginning of the 20th century are collected in Table 5.1.

Table 5.1: Past Pandemics and their Death Toll

Pandemic	Outbreak Year	Death Toll (M)	Death Toll (%)[11]
'Spanish' influenza	1918	50.0 [12]	0.950
'Asian' influenza	1957	1.1 [13]	0.038
'Hong Kong' influenza	1968	1.0 [14]	0.028
HIV/AIDS	1981	$\geq$ 38.0 [15]	$\geq$ 0.840
'Swine' influenza	2009	0.3 [16]	0.004
COVID-19	2019	$\geq$ 4.7 [17]	$\geq$ 0.060

This table reports the pandemics of the 20th and 21st centuries, and their death toll (both in millions, and as a percentage of the global population at the time).

To calibrate pandemic frequency it would be inappropriate to take a value of 0.06 (representing the 6 pandemics in the last 100 years) since there is likely selection bias in this calibration as the time period in question begins and concludes with a pandemic. Therefore I calibrate pandemic frequency to the 4 pandemics observed over the 20th century, yielding $\lambda = 0.04$. I also consider a higher calibration of

[11] Here I take the percentage of the total global population at the time of the outbreak using UN population data from 1950-2020 (United Nations Population Division, 2019), and estimates from Roser, Ritchie, and Ortiz-Ospina (2013) for the global population earlier than 1950. Note that this method will over-estimate the death toll percentage for outbreaks that extend over many decades such as HIV/AIDS since the global population increases over this time. However, this will be somewhat balanced by the fact that many individuals are still dying from AIDS today and will into the future, at a rate of around 700,000 a year (United Nations, 2020).

[12] (Taubenberger and Morens, 2006).

[13] (Viboud et al., 2016).

[14] (Kilbourne, 2006).

[15] HIV/AIDs was first recognised as an epidemic in 1981, however the virus is believed to have emerged around fifty years earlier (Greene, 2007; Sharp and Hahn, 2011). The death toll is taken from United Nations (2020).

[16] (Dawood et al., 2012).

[17] This is the very lowest bound death toll: the number of confirmed deaths to October 2021 (Roser et al., 2020). The true total death toll, calculated by excess mortality, is likely to be 1.3 to 2.2 times higher and will not be certain until the conclusion of the pandemic (Sanmarchi et al., 2021).

$\lambda = 0.06$, since, as I have already suggested in subsection 2.1.2, we are likely to see an increase in the rate of pandemics relative to the past.

To calibrate the death toll from pandemics, I take the average of the mortalities from the 20th-century pandemics, equating to roughly .46% of the population dying from the disease,[18] corresponding to an impact parameter of $\beta_d = 217.4$.[19] One problem with the model using death toll is that this metric is insensitive to lost life years. This is problematic since not all pandemics have the same age-mortality distribution.[20] Though, the main focus of this chapter is to understand how considering existential risk affects WTP for pandemic mitigation activities. A more rigorous measure of the loss of human life from pandemics may improve the accuracy of the contribution of lost lives to WTP to mitigate pandemics, however it will not interact in a significant way with the introduction of an endogenous discount factor capturing a change in existential risk. Therefore, I proceed by calibrating via death toll despite this limitation.

To calibrate the expected (non-death) costs to the global economy of pandemics I refer to estimates from McKibbin and Sidorenko (2006), where the economic burden is calculated using a dynamic model following a pandemic shock. The authors consider several calibrations of the size of the pandemic shock, including a 'mild' pandemic resulting in 1.4 million deaths, a 'moderate' pandemic resulting in 14 million deaths, a 'severe' pandemic resulting in 71 million deaths, and an 'ultra' pandemic resulting in 142 million deaths. The resulting economic consequences from these calibrations are a .7%, 2.8%, 7% and 13.6% loss in GDP respectively.[21] I will assume that the expected cost is associated with the moderate pandemic, in which case the corresponding impact parameter is $\beta_c = 35.7$.[22] For illustration, in Appendix B, I also consider the case where expected cost is equal to that of a mild pandemic; under this calibration absolute WTP will change, though the change in WTP induced by existential considerations is roughly the same.

To calibrate the extinction probability of pandemics, δ_p, I consider a range of

[18] In Martin and Pindyck (2021), the death rate from pandemics is calibrated to the US mortality rate from the Spanish Flu of 4%. The authors offer little justification for this calibration, and we ought to wonder why it should be calibrated to the most severe global pandemic in a single country. However, this difference in calibration is partially compensated by the fact that the authors choose a lower mean arrival frequency of $\lambda = 0.02$ than I have applied here. Regardless, it seems their analysis is calibrated with little regard to empirical evidence, and without justification for their approach.

[19] $\beta_d = \frac{1}{\%Population\ Loss}$

[20] For example, the Spanish influenza was notable for its high mortality in the 20-40 age bracket relative to other pandemics (Gagnon et al., 2013).

[21] Given that the COVID-19 pandemic seems to resemble the moderate pandemic most closely by excess mortality, and in 2020 we saw a 7% drop in global GDP compared to forecasted global GDP for 2020 (IMF, 2021), it seems likely that these estimates of economic damage could be underestimated.

[22] $\beta_c = \frac{1}{\%GDP\ loss}$

estimates. The probability of extinction from a pandemic is clearly very uncertain. Therefore, one should interpret the results from these calibrations as a numerical experiment rather than an objective analysis. The highest calibration I consider is that proposed by Ord (2020), so that the annual risk of extinction from pandemics is the sum of 1 in 1,000,000 and 1 in 3,000 corresponding to the risk estimates from naturally occurring and engineered pandemics. Due to the insignificance of the hazard rate from naturally occurring pandemics relative to the hazard rate from engineered pandemics, for simplicity I set the highest calibration at 1 in 3,000.[23] The lowest calibration is set at 1 in 1,000,000 annual chance of existential collapse from pandemics, corresponding to just the natural rate in Ord (assuming that engineered pandemics pose no threat). I also consider a mid-range estimate at the mean of the log of the high and low calibrations, at roughly 1 in 55,000 chance of extinction from pandemics. Given that to arrive at an estimate for the extinction probability from pandemics one must multiply the probability of a pandemic (λ) by the probability of extinction given a pandemic occurs ($P(E|P)$). I further assume that to arrive at Ord's estimate for existential risk from pandemics, then δ_p is derived using the higher calibration for λ of 0.06 than the historical rate of 0.04 since Ord is explicitly concerned about an *increased* risk of pandemics in the future relative to the past. Therefore, dividing each calibration for δ_p by 0.06 yields the $P(E|P)$ for each δ_p: for $\delta_p = 1/3,000$ we have $P(E|P) = 1/180$; for $\delta_p = 1/55,000$ we have $P(E|P) = 1/3,300$; and for $\delta_p = 1/1,000,000$ we have $P(E|P) = 1/60,000$.

As in previous sections, I also apply the following calibrations: $\delta = 0.001$, $g = 0.02$, and $\eta \in [1.5, 5]$.

5.2.2 Population Growth

Until now, for simplicity I have assumed zero population growth. However, there are two reasons population growth considerations are particularly important in the models developed above. Firstly, the model assumes total utilitarianism, meaning that all else the same, multiplying a population size by x will result in the welfare x times higher. Secondly, in an endogenous discount model (or any model that determines δ by evaluating the probability of extinction) the population growth rate has a particularly dramatic effect on the value of future welfare since δ is relatively close to zero. To illustrate this point consider the following numerical example.

The effective SDR applied to future consumption in the above model is given

[23] It appears here that the Ord estimates depart significantly from the historical examples reported in Table 5.1. This is a consequence of Ord's estimates being of annual existential risk from a pandemic rather than the probability of a pandemic, λ. Therefore these estimates do not imply that Ord believes that the annual probability of an engineered pandemic is significantly higher than the probability of a naturally occurring pandemic, just that engineered pandemics are significantly more likely to result in an existential catastrophe.

by:

$$\rho = \delta - n + g(\eta - 1)$$

with δ representing the hazard rate of existential collapse which is calibrated to $\delta = 0.001$. If $g = 0.02$ and $\eta = 1.5$, then changing n from 0 to 0.01 (i.e., changing the population growth rate from 0 to 1%) decreases the discount rate by 91%! However, if δ is not derived using extinction discounting but instead through market related activity (which typically results in $\delta = 0.02$), then changing n from 0 to 0.01 decreases the discount rate by only 33%. Since a low SDR increases the contribution of future generations to evaluated social welfare, we can expect that including sustained population growth of even just 1% within an extinction discounting framework will have a dramatic effect on the resulting evaluation of WTP.

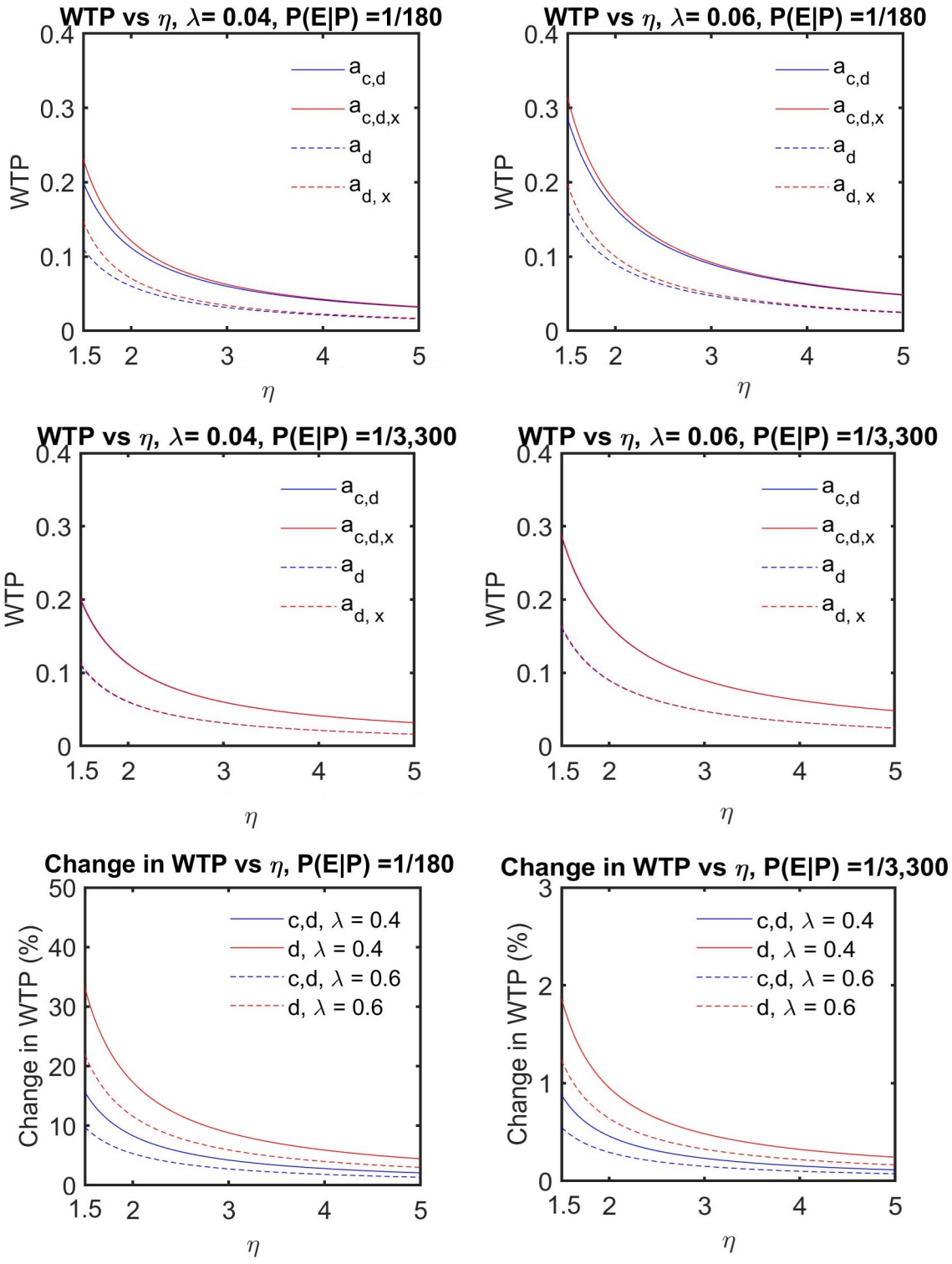
Given that recent demographic research has indicated that the annual global population growth rate will reach zero by the end of the century, assuming either $n = 0$ or $n = 0.01$ is likely to be a mistake since the true population trajectory lies somewhere in between these two values (Bricker and Ibbitson, 2019; United Nations, 2019). However, since this thesis only aims to understand which assumptions lead to significant WTP for mitigation of existential risk, and in that sense is a numerical experiment rather than a precise estimation, I proceed with these two calibrations for population growth. Furthermore, since to my knowledge, population growth has not been considered in a numerical experiment applying the endogenous discount model, the insights from the relationship between population growth and extinction discounting will be novel.

5.3 Results

Applying the above calibrations to equations (5.14) to (5.17) yields the results presented in Figure 5.1 and Figure 5.2. A calibration of $\delta_p = 1/1,000,000$ ($P(E|P) = 1/60,000$) is left out of these figures since it returns negligible results – at this level of existential risk, existential considerations do not affect our WTP to mitigate pandemic risk.

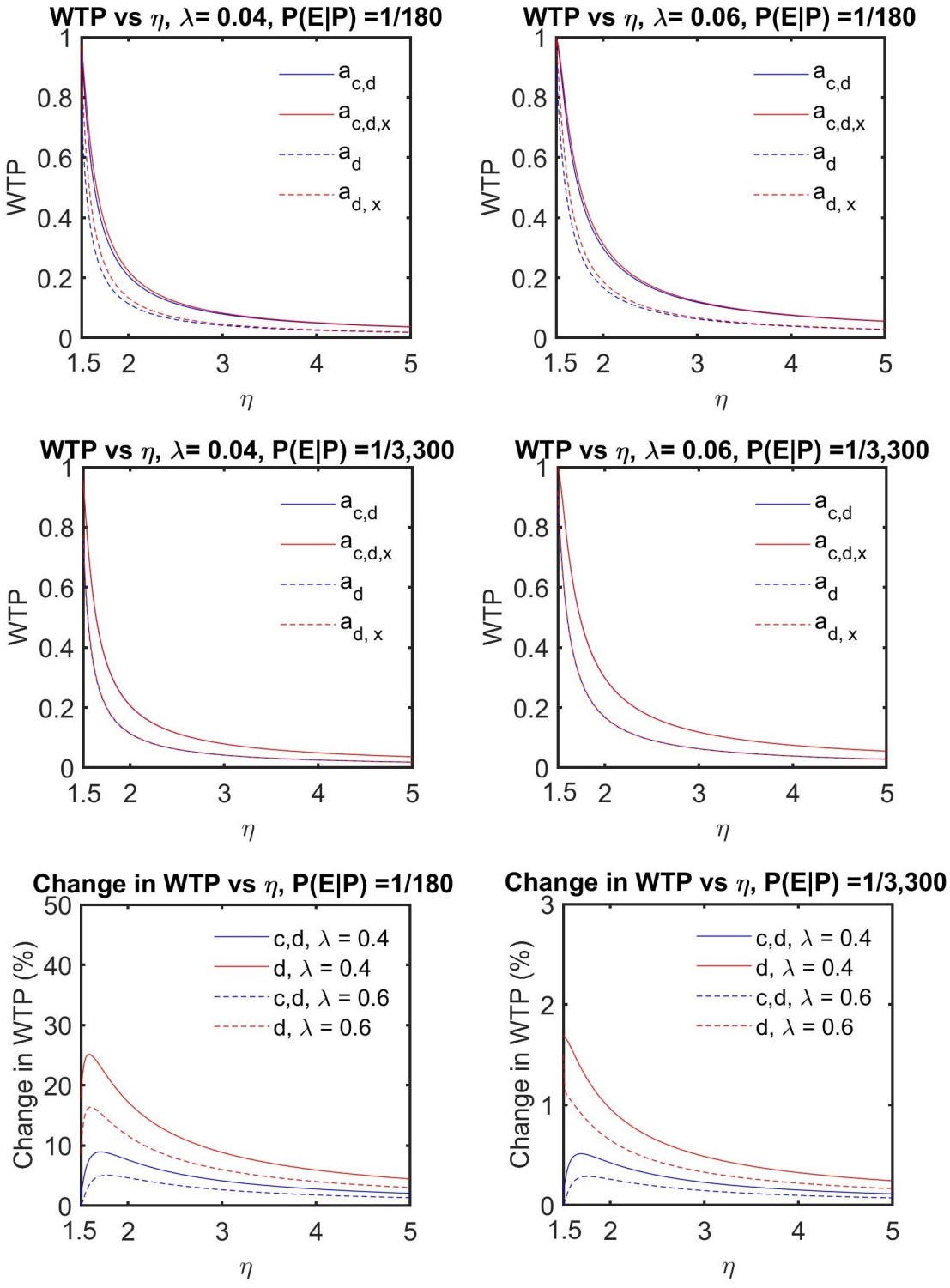
Comparing the solid lines with the dashed lines of the same colour in both figures, we can see that WTP for pandemic mitigation is increased when we consider pandemics to be a threat to both lives and consumption, relative to just considering them a threat to lives as in Martin and Pindyck (2021). This difference is significant, with WTP roughly doubling after including consumption considerations across all values of η. Furthermore, as in the results from section 3.1, these figures indicate that an increase in the annual probability of a pandemic, λ, produces an approximately linear increase in the WTP. For example, going from $P(E|P) = 1/180$ to $P(E|P) =$

Figure 5.1: Willingness to Pay to Avert Pandemics when $n = 0$



This figure presents WTP as a proportion of global income to mitigate all pandemic risk against aversion to intergenerational inequality (η) with no population growth ($n = 0$). The figure considers two calibrations for the annual probability of pandemic ($\lambda = 0.04$ and $\lambda = 0.06$) and two calibrations for the extinction risk given that a pandemic occurs ($P(E|P) = 1/180$ and $P(E|P) = 1/3,300$).

Figure 5.2: Willingness to Pay to Avert Pandemics when $n = 0.01$



This figure presents WTP as a proportion of global income to mitigate all pandemic risk against aversion to intergenerational inequality (η) with population growing by 1% annually ($n = 0.01$). The figure considers two calibrations for the annual probability of pandemic ($\lambda = 0.04$ and $\lambda = 0.06$) and two calibrations for the extinction risk given that a pandemic occurs ($P(E|P) = 1/180$ and $P(E|P) = 1/3,300$).

$1/3,300$ which is approximately an 18-fold decrease in existential risk given that a pandemic occurs, results a roughly 18-fold decrease in WTP, as demonstrated in the bottom two graphs in Figure 5.1 and Figure 5.2. Furthermore, in both Figure 5.1 and Figure 5.2 one can see that the change in WTP as a result of including existential considerations decreases quite significantly as η increases. This finding is consistent with the findings of both section 3.1 and when a similar model is applied to climate change, as in Méjean et al. (2020).

Comparing Figure 5.1 and Figure 5.2, one can see a stark difference in the absolute WTP (rather than the change in WTP). This difference is particularly pronounced at low values of η where WTP becomes close to 1 (though does not reach 1 on any graph). As discussed in the previous section, this is the result of the discount rate being very close to 0 at low values of η when $n = 0.01$. Comparing the change in WTP between $n = 0$ and $n = 0.01$ calibrations, the figures seem to suggest that there is very little interaction between population growth and the introduction of existential risk; this is particularly puzzling since section 3.1 demonstrated that WTP to mitigate a pure existential threat was dramatically increased with the inclusion of population growth – increasing WTP ten-fold at low values of η. The issue with making this comparison is that under population growth, WTP for total pandemic mitigation without considering existential risk ($a_{c,d}$ and a_d) is already *very* large, and particularly so at low values of η. If the social planner were able to allocate most of the total resources to mitigate pandemics then the marginal utility from consumption would be significantly higher given the curvature of the utility function. At low levels of consumption, the marginal utility gains from further spending on mitigating pandemics – by including existential risk as a possible consequence of pandemics – has to be significantly higher to justify diverting further resources away from consumption. Therefore, direct comparison of the change in WTP between Figure 5.1 and Figure 5.2 is very difficult.

5.4 Discussion

Out of the results presented in Figure 5.1 and Figure 5.2 the calibration of $\lambda = 0.06$ and $P(E|P) = 1/180$ has received the greatest support from the literature: annual pandemic probability of 0.06 is slightly higher than the historical rate of 0.04 (as predicted by Jones et al. (2008)), and at this annual probability $P(E|P) = 1/180$ corresponds to Ord's best-guess estimate of existential risk from pandemics of an annual chance of $1/3,000$. Furthermore, zero population growth aligns more closely with the expected rate of population growth for the future United Nations (2019), and low values of η (in the range of 1.5 to 2) are most commonly applied by governments (Harrison, 2010). Under this preferred calibration (of

$P(E|P)$, η, n, and λ), we see that the change in WTP after the inclusion of existential considerations is between roughly 6 and 10% (corresponding to the dashed blue line in the bottom left graph of Figure 5.1). The impact of existential risk is revised upward when assuming a lower pandemic probability (λ) or assuming that pandemics only present a death loss (rather than a consumption and death loss) as demonstrated in the same graph.

As outlined in subsection 5.2.2, the results presented in Figure 5.2 rely on a *very* strong assumption about the trajectory of population growth – that in the periods with out a pandemic shock, the population will continue to grow endlessly at 1% until we encounter an extinction event. While the current global population growth rate is at approximately 1%, this rate has been trending downwards since 1964 and is expected to be roughly 0% by the end of the century (United Nations, 2019). For simplicity, this model has assumed an exogenous population growth rate which has provided novel insights in the context of the application of the endogenous discount method to pandemics, but further research is needed to identify how WTP for pandemic mitigation responds to the application of the expected population trajectory.

These results all share the same general pattern as the results of both Chapter 3 and Chapter 4. Namely, aversion to intergenerational inequality, and population growth have dramatic implications for WTP. However, it is noteworthy that while the pattern of results is the same, the scale of results seems to be very different when comparing the affect of existential risk considerations on a specific intervention to this affect on the WTP to mitigate all pandemic threat. For example, in Chapter 4 it was demonstrated that including existential collapse as a potential consequence of pandemics could decrease the break-even probability of a pandemic intervention by up to 50%, whereas here I have suggested the most likely affect on WTP for total pandemic mitigation is 6-10% (though this affect reaches up to 30% under alternative calibrations). This discrepancy is largely because of the difference in level of total investment to reduce pandemic risk, making it difficult to compare the change in WTP to the change in the break-even probability (just as it is difficult to compare the change in WTP between calibrations of 0 and 1% population growth). For example, the intervention considered in Chapter 4 was suggested to cost in the range of 0.02-0.04% of global GDP, however the WTP to mitigate all pandemic threat (before considering existential risk) was in the range of 10-30% of global GDP. With this significant difference in initial spending, the trade-off between consumption and further spending after introducing existential risk is very different due to difference in the marginal utility of consumption. This comparison is made even more difficult since the results from Chapter 4 consider only COVID-19 scale pandemics whereas

the results presented here are calibrated to all pandemics,[24] and because Chapter 4 includes the costs of the intervention as an absolute value (since is considers a specific intervention with a defined cost) rather than a proportion of income as applied here.

Since the focus of this chapter is how including existential threats as a possible consequence of pandemics can change WTP to mitigate all pandemic threat, the practical applications of these results are, admittedly, limited as the possibility of mitigating *all* pandemic threat seems far-fetched. However, these results do have several important implications for the analysts focusing on evaluating WTP for catastrophic threats. Firstly, under extinction discounting defining population growth accurately is particularly important. For example, in Martin and Pindyck (2021) where extinction discounting is not applied, changing between population growth of zero and 1% only amount to a difference in WTP of about 10% of GDP. However, in the present analysis which does employ extinction discounting, this same change in population growth changes total WTP significantly, and particularly so at low calibrations of aversion to intergenerational inequality. Secondly, including existential risk can still have dramatic impacts on spending decisions, even when the WTP to mitigate pandemic threats without the possibility of existential risk is very high. This serves as a robustness check on the importance of existential risk considerations; even if a government already spends a large amount on pandemic preparedness, if they have ignored the possibility of a pandemic induced existential collapse, then they are likely under-spending by a non-negligible amount. Finally, this chapter further demonstrates the conditions at which ignoring the possibility of existential risk has a very minimal impact on spending decisions related to pandemics. Namely, either having η relatively high, or the probability of existential collapse from a pandemic is very low (here, in the order of a 1/55,000 annual chance) are sufficient to reduce the impact of including existential considerations to a negligible result.

[24] This is evident from the fact that the calibrations applied in Chapter 4 for annual pandemic probability (without any mitigation) were 0.01 and 0.005, whereas in this chapter they are 0.04 and 0.06.

Chapter 6
Discussion and Limitations

6.1 A Bang or a Whimper?

On consequential assumption in any analysis that applies an endogenous discount factor to represent existential risk, is that if a catastrophic event were to result in an existential collapse, then this collapse would happen immediately: a 'bang' scenario. This assumption is evident in that extinction discounting is only sensitive to a binary outcome – extinct or not – an assumption that is difficult to square with the reality of many existential collapse scenarios. For example, in an existential collapse from a pandemic, we might expect that there is some period over which social welfare is decreasing as the pandemic spreads, rather than an overnight species extinction; or in the example existential collapse from climate change, the welfare path is overwhelmingly likely to be one of gradual decline to zero as the planet becomes an increasingly inhospitable place. One can imagine similar arguments for other risks in the risk landscape outlined in section 2.1. Rather than the world ending with a bang, I call these more plausible outcomes 'whimper' scenarios.[1]

One further issue with the endogenous discounting approach is that it assumes that welfare under an existential collapse is zero. Recalling Bostrom's definition of an existential threat introduced in Chapter 1 – the class of catastrophe that could result in human extinction *or* permanently and dramatically reduced human welfare – the endogenous discounting approach may over-estimate the welfare loss associated with an existential collapse, which could be greater than zero under this definition. As suggested in subsection 2.2.2, there is no convenient definition of exactly how close to zero the welfare loss induced by a catastrophic event must be to classify it is an existential collapse. Therefore, there is no easy way to evaluate the resulting bias in models that employ an endogenous discount factor. One solution to this issue is to simply calibrate the models employed above to one's preferred estimate of the probability that an existential risk will reduce welfare to exactly zero, rather than calibrating to any existential threat (which include zero welfare and near-zero welfare scenarios). However, under this approach one would then have to develop a satisfactory method of accounting to include the welfare consequences

[1] I take these labels from T.S. Elliot's poem *The Hollow Men* (1925): 'This is the way the world ends: Not with a bang but a whimper'.

from near-zero welfare existential collapses which maybe be immensely difficult given the uncertainty around such scenarios.[2]

Given the difficulties of an alternative approach, and despite the limitations of endogenous discount model, the endogenous discount approach currently seems to be the best method of evaluating changes in existential risk despite the limitations of the approach for two reasons. Firstly, it avoids the need to predict the exact social welfare path both before and after an existential collapse. Given our inherent lack of experience with existential collapse, predicting this path is very dubious. Secondly, if we expect that the welfare path is different across various existential threats, then by assuming a bang scenario the models applied remain general and can be applied to the WTP analysis of any existential threat.

6.2 Benefits and Limits of Willingness to Pay

The analysis presented here has applied WTP to understand how existential considerations can affect decisions regarding resource allocation. WTP is particularly useful in this setting because it avoids the need for a specified functional form of the relationship between safety spending and hazard rate – a relationship that is impossible to define empirically. Instead, with WTP analysis it is sufficient to know welfare conditions under a status quo and alternate path; furthermore, the alternate path does not, in principle, have to be some realistic scenario (such as the complete mitigation of all pandemic risk as considered in Chapter 5).

A further advantage of WTP is that it does not require the same complete understanding of the opportunity costs of investing in pandemic mitigation that is required to evaluate optimal spending, since optimal resource allocation is determined by the marginal benefit of all spending opportunities. In WTP analysis it is perfectly valid to ignore other spending opportunities since the only requirement is that under some new policy or investment, the new social welfare is equal to the status quo welfare. One unfortunate consequence of this weak requirement for WTP analysis is that a positive WTP for a particular project does not imply that a social planner should even spend any resources at all on that project. For example, in Chapter 3 I demonstrated that, over the calibrations considered, a social planner is willing to spend up to 35% of consumption to mitigate up to 20% of existential risk, though it could be also true that there always exist other spending opportunities with greater marginal benefit. In this case, it is optimal for the social planner to dedicate no extra resources to risk mitigation, even if WTP was still significantly different from zero.

[2]For example, if the human population was reduced by 99.9% from a catastrophe, it is very difficult to anticipate the ensuing path of human welfare.

Since a WTP analysis cannot derive optimal conditions, this thesis offers little in the way of precise spending recommendations. However, given that this analysis aims to provide a sense of scale of the effect of existential considerations on the estimates of the seriousness of pandemic risk, the WTP approach is appropriate in this context.

6.3 Implications for Pandemic Preparedness Policy

A major finding of this thesis is that under certain assumptions – low aversion to intergenerational equality combined with best-guess estimates of the probability of existential collapse from pandemics – the effect of existential considerations on resulting WTP or cost-effective estimates related to pandemics is non-negligible. This finding is important since these considerations have been ignored by virtually all cost-benefit analyses of pandemic risk mitigation interventions.[3] The major problem with these findings is that the calibrations of key variables which they are derived from are highly uncertain, and uncertain in several ways.

When it comes to calibrating aversion to intergenerational inequality, there is *methodological* uncertainty about how we should derive this value (be it normative or positive). There also is *moral* uncertainty, such that even if we agreed on a normative approach then we still don't know exactly what this value should take.[4] When it comes to calibrating the probability of existential risk, due to the very limited research on the topic I rely entirely on Ord (2020) estimates to suggest a favoured calibration in the numerical experiments carried out in Chapters 4 and 5. Since any research into estimating existential risk is necessarily highly subjective, reducing uncertainty is difficult. Though one feasible approach to reducing *some* uncertainty related to these estimates is to expand the set of estimates from which we can draw on to conduct numerical experiments through funding research into existential risk. Given the very small pool of existing research and the demonstrated WTP for existential risk mitigation, it seems that this research would be immensely valuable for understanding intergenerational resource allocation.

This analysis also demonstrated the degree to which broadening the risk mitigation effects of an intervention can dramatically increase its cost-effectiveness. For example, this thesis has shown that the cost-effectiveness of an intervention to mitigate the threat of naturally arising pandemics is highly dependent on the extent to which it also mitigates the threat of engineered pandemics. I have

[3] With the exception of Millet and Snyder-Beattie (2017).

[4] If economists agreed on a positive approach, this level of uncertainty would be diminished somewhat since the value can be inferred empirically through data on interest rates and growth as in Weitzman (2007). See MacAskill, Bykvist, and Ord (2020) for a detailed discussion of moral uncertainty.

noted that many pandemic interventions are suggested to have this broad reach. For example, achieving broach commitment to international health regulations of the World Health Organization (WHO) – 'One Health' – is claimed by WHO to reduce pandemic risk regardless of the origin of the outbreak. Not only does this thesis suggest that the cost-effectiveness of interventions like One Health might be dramatically underestimated, it also suggests that there is a case for expanding existing interventions to target mitigating some level of existential risk if they do not do so currently. For example, the pandemic mitigation portfolio proposed in Dobson et al. (2020) may be significantly more cost-effective if it could also include interventions directly targeted at effective means of reducing engineered pandemic risk.

Chapter 7
Conclusion

Through the application of an endogenous discount model, this thesis has demonstrated the necessary conditions for a social planner to be willing to spend a non-negligible proportion of global income on existential risk mitigation. The exact willingness to pay for existential risk mitigation is highly dependent on key variables, including aversion to intergenerational inequality, population growth, and the effectiveness of the intervention at mitigating risk. Previous work has demonstrated that some of these variables have a similar result on optimal climate policy under the possibility of existential collapse from climate change, but this novel approach to understanding the impact of existential risk on spending decisions presents the importance of these variables in a model that can be applied to any existential threat.

In a novel application of the endogenous discount method, this thesis has also demonstrated that the cost-effectiveness of a specific intervention to mitigate pandemic risk can be significantly increased after including a change in total existential risk (by up to 50%). Again, this change is dependent on the calibration applied. In particular, the amount of existential risk that is mitigated by the intervention has a dramatic impact on the cost-effectiveness. Unfortunately there is no empirical method of clarifying the key uncertainties related to the impact of the intervention on existential risk, so we must necessarily rely on best-guess estimates. Further research could refine these estimates, leading to more robust results for the impact of existential considerations on spending decisions related to pandemics than presented in this thesis.

This thesis also considers the impact of existential risk in a dynamic model of willingness to pay that considers the social welfare effects of population and consumption loss. This analysis demonstrates that under what I consider to be the calibration of key variables most strongly supported by the literature, willingness to pay increases by 6-10% after the introduction of existential risk, with many other calibrations increasing willingness to pay by even more.

Several simplifying assumptions were made in the modelling applied in this thesis. Firstly, I have assumed that economic growth is exogenously determined, and constant. If the social planner could invest in growth, instead of just consumption or risk mitigation, then WTP would be less than suggested in Chapters 3, 4 and 5.

However, I expect this effect to be small, particularly in the results from Chapter 3 (when there is no population growth) and 4, since the amount of total investment in the projects being discuses is a very small proportion of total income. Secondly, I have assumed either zero population growth or continuous annual population growth of 1%. This is at odds with recent population research that has suggested the global population growth rate will reach zero by the end of the century. Therefore, the true WTP for risk mitigation activity will likely lie somewhere between the two estimates of WTP under zero and 1% annual growth. Thirdly, I have assumed that the only contributor to social welfare is individual consumption. Though, as this thesis is focused on how a social planner should trade-off between current resources and species longevity, limiting social welfare to consumption of resources captures the problem adequately.

Due to the *significant* uncertainty around many parameters considered here, I have opted for a numerical experiment rather than suggesting that some single calibration exactly represents the willingness to pay, or break-even probability for some pandemic intervention. However, where possible I have also indicated which calibrations seem to have the most significant support from the literature on pandemics and catastrophic risks. Under these calibrations, existential considerations have a non-negligible impact on the analysis of spending decisions. In conclusion, a policymaker ought to be clear about their underlying assumptions related to existential risk, failing to do so can significantly bias estimates of the cost-effectiveness of risk mitigation projects.

Appendix A
Derivations and Modeling

A.1 The Social Discount Rate

This thesis applies four different forms of the social discount rates (SDRs) which are reiterated below. I also define another SDR formally here.

Chapter 2 cites the SDR derived in Ramsey (1928):

$$\rho = \delta + g\eta \tag{A.1}$$

Chapter 3 defines two effective SDRs, one with population growth, n, and one without:

$$1 + \rho \equiv (1 + \delta)(1 + g)^{\eta-1}(1 + n)^{-1} \tag{A.2}$$

$$1 + \rho \equiv (1 + \delta)(1 + g)^{\eta-1} \tag{A.3}$$

and Chapter 5 defines the effective SDR also with population growth applied in Martin and Pindyck (2021):

$$\rho \equiv \delta - n + g(\eta - 1) \tag{A.4}$$

These four SDRs differ across three dimensions: whether population growth is included, whether they are derived from a social welfare function or defined for the purpose of simplification,[1] and whether they correspond to a social welfare function defined in discrete or continuous time. Before offering a direct comparison of these SDRs, I derive an SDR missing from the above collection; importantly, this SDR is derived in discrete time, making it the formal SDR for the social welfare functions applied in Chapters 3 and 4.

Definition 1. *Taking a social welfare function defined in discrete time:*

$$W = \sum_{t=0}^{\infty} \frac{1}{(1+\delta)^t} N_t \frac{c_t^{1-\eta}}{1-\eta}$$

[1] I call these 'effective' SDRs.

where $c_t = c_0(1+g)^t$ *and* $N_t = N_0(1+n)^t$. *The social discount rate,* ρ, *is:*

$$1 + \rho = \left(\frac{\partial W_0 / \partial c_0}{\partial W_0 / \partial c_t} \right)^{\frac{1}{t}} = (1+\delta)(1+g)^{\eta}(1+n)^{-1} \tag{A.5}$$

Table A.1 summarises the key differences between the five SDRs described above.

Table A.1: Comparison of Various Expressions for the Social Discount Rate

SDR	Population Growth	Derived	Discrete Time
A.1	No	Yes	No
A.2	Yes	No	Yes
A.3	No	No	Yes
A.4	Yes	No	No
A.5	Yes	Yes	Yes

This table illustrates the key differences between five different SDRs detailed in this section along three dimensions: if population growth is included ('Population Growth'), if it is derived from a social welfare function rather than defined for the purpose of simplification ('Derived'), and if it corresponds to a social welfare function defined in discrete rather than continuous time ('Discrete Time').

There is an obvious difference in the form of the expression for the SDR in continuous compared to discrete time, with terms entering additively in the continuous case and multiplicatively in the continuous case. However, as noted in Dasgupta (2008), SDRs derived from discrete welfare functions will be approximately equal to the SDRs derived from continuous welfare functions under the range of values applied for δ, n, g, and η.

A.2 Derivation of Willingness To Pay to Mitigate Existential threats

Starting with the social welfare functions under the status quo, W_0, and under an investment in a risk mitigation project W_1:

$$\begin{aligned} W_0 &= \sum_{t=0}^{\infty} \theta_0^t \times N_0(1+n)^t \times \frac{(c_0(1+g)^t)^{1-\eta}}{1-\eta} \\ W_1 &= \sum_{t=0}^{\infty} \theta_1(x)^t \times N_0(1+n)^t \times \frac{((1-a)c_0(1+g)^t)^{1-\eta}}{1-\eta} \end{aligned}$$

Applying the formula for a geometric series we have

$$\begin{aligned} W_0 &= N_0 \times D_0 \times U_0 \\ W_1 &= N_0 \times D_1 \times U_1 \end{aligned}$$

where

$$\begin{aligned} D_0 &= \frac{1}{1-\theta_0(1+g)^{1-\eta}(1+n)} \\ D_1 &= \frac{1}{1-\theta_1(x)(1+g)^{1-\eta}(1+n)} \\ U_0 &= \frac{c_0^{1-\eta}}{1-\eta} \\ U_1 &= \frac{((1-a)c_0)^{1-\eta}}{1-\eta} \end{aligned}$$

As in section 3.1 (and A.2 above), we can define two SDRs:

$$\begin{aligned} 1+\rho_0 &\equiv (1+\delta)(1+g)^{\eta-1}(1+n)^{-1} \implies D_0 = \frac{1}{1-(1+\rho_0)^{-1}} \\ 1+\rho_1(x) &\equiv (1+(1-x)\delta)(1+g)^{\eta-1}(1+n)^{-1} \implies D_1 = \frac{1}{1-(1+\rho_1(x))^{-1}} \end{aligned}$$

If $|(1+\rho_0)^{-1}| \geq 1$ or $|(1+\rho_1(x))^{-1}| \geq 1$ then the geometric formula cannot be applied. Intuitively values of ρ in this range imply that the value of the future is undefined and hence trade-offs can not be considered in this range. Therefore, I only consider values calibrations such that $|(1+\rho_0)^{-1}| < 1$ or $|(1+\rho_1(x))^{-1}| < 1$. Allowing $W_0 = W_1$ yields:

$$\begin{aligned} & D_0 \times U_0 = D_1 \times U_1 \\ \implies & \frac{U_1 - U_0}{U_0} = -\frac{D_1 - D_0}{D_1} \end{aligned}$$

Where $\frac{U_1-U_0}{U_0}$ is the percentage change in U relative to U_0, and when U is positive and $\frac{D_1-D_0}{D_0}$ is the percentage change in D relative to D_1. In section 3.1 I note that negative utility functions can yield problematic results when comparing welfare with endogenous discounting, and by taking the change in the utility this problem can be avoided, where the percentage change in any negative number (in this case U_0 and U_1) is equal to $\frac{U_1-U_0}{|U_0|}$. Therefore, the above equation becomes:

$$\frac{U_1 - U_0}{|U_0|} = -\frac{D_1 - D_0}{D_1}$$

and since over all calibrations considered in this thesis utility is negative, then this we have:

$$-\frac{U_1 - U_0}{U_0} = -\frac{D_1 - D_0}{D_1}$$

Substituting in U_0, U_1, D_0, and D_1 yields:

$$\frac{\frac{((1-a)c_0)^{1-\eta}}{1-\eta}}{\frac{c_0^{1-\eta}}{1-\eta}} - 1 = 1 - \frac{1-(1+\rho_1(x))^{-1}}{1-(1+\rho_0)^{-1}}$$

And solving for a yields:

$$a = 1 - \left[2 - \frac{1-(1+\rho_1(x))^{-1}}{1-(1+\rho_0)^{-1}}\right]^{\frac{1}{1-\eta}}$$

Giving the result presented in section 3.1. Note that one can also derive a solution which is approximately equal to the above solution over the values for θ, η, and g considered:

$$a \approx \hat{a} = 1 - \left[\frac{1-(1+\rho_0)^{-1}}{1-(1+\rho_1(x))^{-1}}\right]^{\frac{1}{1-\eta}}$$

Where $\hat{a}$ is derived by allowing $\frac{U_1-U_0}{|U_0|} = -\frac{D_1-D_0}{D_0}$. Numerically, one can see that $\hat{a}$ does not deviate from a by more than a couple of percent of a over the range of values considered.

A.3 Proof of Proposition 3.3.1.

Proposition 3.3.1. *With the social welfare function defined by W:*

$$W = \sum_{t=0}^{\infty}[\theta(a)]^t \times N \times \left[\frac{((1-a)c_0(1+g)^t)^{1-\eta}}{1-\eta}\right]$$

optimal a is strictly increasing in c_0 if

$$\theta(a) > 1 - \frac{\delta_0}{1+\delta_0}(1-a)^{\eta-1}$$

Proof. Optimal a is strictly increasing in c_0 implies that the marginal welfare gains from safety spending increase relative to marginal welfare gains from consumption as consumption strictly increases, i.e., $\frac{\partial W/\partial a}{\partial W/\partial c_0}$ strictly increases in c_0. Therefore, we need:

$$\frac{\partial \frac{\partial W/\partial a}{\partial W/\partial c_0}}{\partial c_0} > 0$$

where

$$\frac{\partial W}{\partial a} = \frac{N\big(c_0(1+g)^t\big)^{1-\eta}}{(1-\theta(a))(1-a)^\eta}\left[\frac{\theta'(a)(1-a)}{(1-\theta(a))(\eta-1)}-1\right] \tag{A.6}$$

$$\frac{\partial W}{\partial c_0} = \frac{N\big((1-a)(1+g)^t\big)^{1-\eta}}{\big(1-\theta(a)\big)c_0^\eta} \tag{A.7}$$

Therefore, dividing (A.6) by (A.7) yields:

$$\frac{\partial W/\partial a}{\partial W/\partial c_0} = \frac{c_0}{(1-a)}\left[\frac{\theta'(a)(1-a)}{(1-\theta(a))(\eta-1)}-1\right] \tag{A.8}$$

And taking the derivative of (A.8) with respect to c_0 yields:

$$\begin{aligned}\frac{\partial \frac{\partial W/\partial a}{\partial W/\partial c_0}}{\partial c_0} &= \frac{\theta'(a)}{(1-\theta(a))(\eta-1)}-\frac{1}{1-a}>0\\ \implies \theta'(a) &> \frac{(1-\theta(a))(\eta-1)}{1-a}\end{aligned} \tag{A.9}$$

Equivalently to (A.9) we have:

$$-\theta'(a) < -\theta(a)\frac{1-\eta}{1-a}+\frac{1-\eta}{1-a} \tag{A.10}$$

Allowing $\phi(a)=-\theta(a)$ and $\phi'(a)=-\theta'(a)$, and substituting into (A.10) yields:

$$\phi'(a) < \phi(a)\frac{1-\eta}{1-a}+\frac{1-\eta}{1-a} \tag{A.11}$$

Since this is a first-order linear differential inequality, it can be simplified further to remove $\theta'(a)$. Gronwall's Inequality Theorem[2] (GIE) in a differential form tells us if

$$u'(t) < a(t)u(t)+b(t) \quad and \quad u(t_0)=u_0$$

Then

$$u(t) < u_0 e^{\int_{t_0}^t a} + \int_{t_0}^t e^{\int_s^t a} b(s)ds$$

Applying GIE to (A.11) yields:

$$\begin{aligned}\phi(a) &< \phi_0 e^{\int_{a_0}^a \frac{1-\eta}{1-a}} + \int_{a_0}^a e^{\int_t^a \frac{1-\eta}{1-a}}\frac{1-\eta}{1-t}dt\\ &= \phi(a)e^{\int_{a_0}^a \frac{1-\eta}{1-a}} + \int_{a_0}^a e^{\int_t^a \frac{1-\eta}{1-a}}\frac{1-\eta}{1-t}dt\end{aligned} \tag{A.12}$$

[2]This form is initially considered in Reid (1930). However, I employ a more commonly applied solution, for example, used in: https://sites.math.washington.edu/~burke/crs/555/555_notes/exist.pdf

Simplifying (A.12) yields:

$$\phi(a) < (\phi_0 + 1)(1-a)^{\eta-1} - 1 \tag{A.13}$$

And substituting in $-\theta(a) = \phi(a)$ and $-\theta_0 = \phi_0$ into (A.13) yields:

$$\begin{aligned} -\theta(a) &< (-\theta_0 + 1)(1-a)^{\eta-1} - 1 \\ \implies \theta(a) &> (\theta_0 - 1)(1-a)^{\eta-1} + 1 \end{aligned} \tag{A.14}$$

And substituting $\theta_0 = \frac{1}{1+\delta_0}$, where $\delta_0 = \delta(a_0)$, into (A.14) yields:

$$\theta(a) > 1 - \frac{\delta_0}{1+\delta_0}(1-a)^{\eta-1} \tag{A.15}$$

□

A.4 Calibrations Conforming to the Restriction in Proposition 3.2.1

To assess which calibrations of intergenerational inequality aversion, η, safety spending as a proportion of total wealth, a, and the discount rate, $\theta(a)$ conform to the restriction, (A.10), from Proposition 3.2.1, first I define an equivalent proposition:

Proposition 3.3.1*. *According to (3.2), optimal a is strictly increasing in c_0 if*

$$\delta(a) < \frac{A}{1-A} \quad \textit{where} \quad A = \frac{\delta_0}{1+\delta_0}(1-a)^{\eta-1}$$

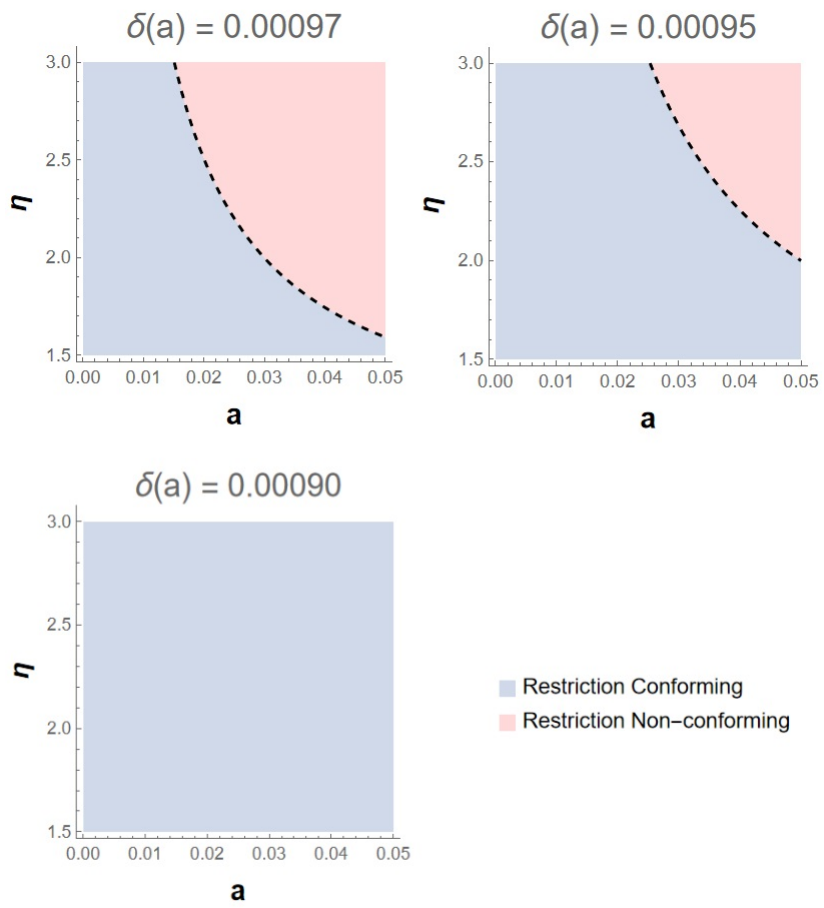
Proof. Substituting $\theta(a) = \frac{1}{1+\delta(a)}$ into (A.10) yields the result. □

Applying this equivalent restriction with the assumption that $\delta_0 = 0.001$, Figure A.1 illustrates acceptable calibrations such that the proportion of optimal safety spending will be increasing with economic growth. This figure limits values to likely candidates for η, a, and $\delta(a)$. I limit the range of η to between 1.5 and 3 following the typical values applied in the literature.[3] I limit the range of a to between 0 and 5% of total wealth. I have chosen this range due to the findings from section 3.1 which suggest that for relatively low changes in the hazard rate as a result of some project, the WTP is at most a few of percent (with no population growth). For example, when $\eta = 1.5$ the social planner has a WTP of roughly 1.8% of total wealth to mitigate 5% of total existential threat. Furthermore, optimal a

[3] See Table 3.1.

is necessarily less than WTP since at the maximum WTP, the welfare maximising social planner is indifferent between the status quo and the spending the maximum WTP on the mitigation project.

Figure A.1: Calibrations Conforming to the Restriction in Proposition 3.3.1



This figure illustrates the values of η, a, and $\delta(a)$ for which the restriction from Proposition 3.2.1 is satisfied, in which case investment in existential risk mitigation will be an increasing proportion of global income.

As we can see from Figure A.1, as a increases this places a stronger restriction on the acceptable values of η so that the calibration is restriction conforming. Comparing across the graphs for different post-investment hazard rates, we can see that if the hazard rate at optimal investment is lower, then this weakens the restriction on η and a. For example, we can see in the bottom left graph of Figure A.1 that if the hazard rate at optimal investment is 10% less after the safety spending, (i.e., $\delta(a) = 0.0009$) then the entire range of likely calibrations for η and a are restriction conforming. By contrast, if the hazard rate at optimal investment is 3% less after the safety spending then as a rule of thumb, either calibration of η

must be high, or safety spending must be very expensive – spending in the order of 3-5% of global GDP to mitigate existential risk by 3%. Given the many spending opportunities for risk mitigation,[4] it seems likely that there are sufficient low-hanging fruit for the responsiveness of existential risk change to be quite large if we were to spend as much as a couple of percent of total GDP on it's mitigation.

A.5 Deriving a Simplified Restriction.

Taking (A.9) from section A.3, we know

$$\theta'(a) \quad > \quad \frac{(1-\theta(a))(\eta-1)}{1-a} \tag{A.16}$$

Since we make the simplifying assumption that $\delta(a) = \delta_0 e^{-\gamma a}$, this implies the following:

$$\theta(a) = \frac{1}{1+\delta_0 e^{-\gamma a}} \tag{A.17}$$

$$1-\theta(a) = \frac{\delta_0 e^{-\gamma a}}{1+\delta_0 e^{-\gamma a}} \tag{A.18}$$

$$\theta'(a) = \frac{\gamma \delta_0 e^{-\gamma a}}{(1+\delta_0 e^{-\gamma a})^2} \tag{A.19}$$

and together (A.16), (A.17), (A.18), and (A.19) imply

$$\theta(a) > \frac{\eta-1}{\gamma(1-a)} \tag{A.20}$$

A.6 Proof of Proposition 3.4.1

Proposition 3.4.1. *If W_0 and W_1 are defined, then $a^* > a$.*

Proof. Suppose $a^* > a$. Then

$$\left[1-\left(2-\frac{1-\big(1+\rho_1(x)\big)^{-1}}{1-(1+\rho_0)^{-1}}\right)^{\frac{1}{1-\eta}}\right]+\alpha\mathbb{E}[d] > \left[1-\left(2-\frac{1-\big(1+\rho_1(x)\big)^{-1}}{1-(1+\rho_0)^{-1}}\right)^{\frac{1}{1-\eta}}\right](1-\mathbb{E}[d])+\alpha\mathbb{E}[d] \tag{A.21}$$

[4] For example, see The Centre for Long-Term Resilience (2021) and Leigh (2021).

where, following section 3.1, I define the effective SDRs:

$$\begin{aligned} 1+\rho_0 &\equiv \frac{(1+\delta)(1+g)^{\eta-1}}{1+n} \\ 1+\rho_1(x) &\equiv \frac{(1+(1+x)\delta)(1+g)^{\eta-1}}{1+n} \end{aligned}$$

W_0 is defined when $|(1+\rho_0)^{-1}| < 1 \implies \rho_0 > 0$ as per the formula for a geometric series. The same holds for W_1, where $\rho_1(x) > 0$. By assumption, $\delta, g, \eta > 0, n \geq 0$ and $x \in (0,1)$, therefore $\rho_0 > \rho_1(x) > 0$. This implies

$$1 > \left(2 - \frac{1-\big(1+\rho_1(x)\big)^{-1}}{1-(1+\rho_0)^{-1}}\right)^{\frac{1}{1-\eta}} > 0$$

and the rearranging (A.21) yields

$$\mathbb{E}[d] > 0$$

and it is assumed that $\mathbb{E}[d] \in (0,1)$. □

A.7 Detail on Dynamic Model of Willingness To Pay to Mitigate Pandemics

This section covers some of the detail of the Martin and Pindyck (2021) model for WTP to mitigate catastrophes that was left out of Chapter 5.

A.7.1 Welfare Contributions of the Dead

In Chapter 5, I note that the model considers how deaths from a catastrophe can still contribute positively to social welfare via bequests. Following Martin and Pindyck (2021), to calculate the welfare contributions of the dead we start with an equation for VSL, defined by an individual's willingness to trade off consumption for a change in probability of death:

$$VSL = \frac{dw}{d(1-p)} = \frac{u(w)-v(w)}{(1-p)u'(w)-pv'(w)}$$

where w can be either the wealth or lifetime consumption of an individual, p is the probability of death, making $(1-p)$ the probability of survival, $u(\cdot)$ describes utility for the living, and $v(\cdot)$ describes utility from the dead where $u(w) > v(w)$ and $u'(w) > v'(w)$.

Since we are considering catastrophes that are very unlikely, Martin and

Pindyck (2021) evaluate VSL at $p = 0$. Therefore, in our application we have:

$$VSL_{p=0} = \frac{u(w) - v(w)}{u'(w)} \tag{A.22}$$

As outlined in Chapter 5, we already have existing estimates for VSL that are frequently applied by governments and policymakers. Here this VSL is considered in the form:

$$VSL = sw \tag{A.23}$$

where s is some positive number that describes how many lifetime incomes a government is willing to pay for a life. For example, at $s = 7$ a the VSL is seven times the lifetime income.

Equating (A.22) and (A.23) yields the welfare contribution from the dead:

$$v(w) = u(w) - swu'(w) \tag{A.24}$$

One can further simplify (A.24) by substituting in the utility function and its derivative; where $u(w) = \frac{w^{1-\eta}}{1-\eta}$, and therefore $u'(w) = w^{-\eta}$. Therefore (A.24) becomes:

$$\begin{aligned} v(w) &= \frac{w^{1-\eta}}{1-\eta} - sw^{1-\eta} \\ &= \frac{1}{1-\eta}\left[w[1 + s(\eta - 1)]^{\frac{1}{1-\eta}}\right]^{1-\eta} \\ &= u(w\varepsilon) \end{aligned}$$

where $\varepsilon = [1 + s(\eta - 1)]^{\frac{1}{1-\eta}}$.

A.7.2 Uniting Existential Risk and Catastrophic Threats

One problem with uniting the endogenous discount model outlined in section 3.1 with the Martin and Pindyck (2021) model of WTP to avert a catastrophe is that they both deal with the problem of negative utility differently. I illustrate this difference with an example below.

Take the status quo welfare function, W_0, and the welfare function after completely mitigating a catastrophe causing death a consumption loss, W_1, from

the Martin and Pindyck (2021) model:[5]

$$\begin{aligned} W_0 &= U_0 \times D_0 \\ W_1 &= U_1 \times D_1 \end{aligned}$$

where, as in section 5.1, we have:

$$\begin{aligned} U_0 &= \frac{1}{1-\eta} \\ U_1 &= \frac{(1-a_{c,d})^{1-\eta}}{1-\eta} \\ D_0 &= \frac{1-\varepsilon^{1-\eta}}{\delta - \boldsymbol{\kappa}_N(1) - \boldsymbol{\kappa}_C(1-\eta)} + \frac{\varepsilon^{1-\eta}}{\delta - \boldsymbol{\kappa}^*_N(1) - \boldsymbol{\kappa}_C(1-\eta)} \\ D_1^{c,d} &= \frac{1}{\delta - \boldsymbol{\kappa}^*_N(1) - \boldsymbol{\kappa}^*_C(1-\eta)} \end{aligned}$$

Since it is assumed that $a_{c,d} > 0$ (reducing risk requires spending), this implies $U_0 > U_1$. Furthermore, when allowing $W_0 = W_1$ to solve for $a_{c,d}$, since $U_1 < U_0 < 0$ then this implies $D_0 > D_1^{c,d}$; and the larger the difference between D_0 and $D_1^{c,d}$, the greater the magnitude of $a_{c,d}$ such that $W_0 = W_1$. Note that if utility were positive we would have $D_0 < D_1^{c,d}$ as was the case in the initial application of the endogenous discount model in section 3.1.

When applying the endogenous discount model we are simply changing δ in D_1 to $\delta_{p\prime}$, where $\delta > \delta_{p\prime}$, yielding:

$$D_1^{c,d,x} = \frac{1}{\delta_{p\prime} - \boldsymbol{\kappa}^*_N(1) - \boldsymbol{\kappa}^*_C(1-\eta)}$$

where equating $U_0 \times D_0$ with $U_1 \times D_1^{c,d,x}$ allows one to solve for $a_{c,d,x}$.

The issue with combining the endogenous discount model with the Marin and Pindyck model is that since $\delta > \delta_{p\prime}$, then $D_1^{c,d} < D_1^{c,d,x} < D_0$; and since the greater the difference between D_0 and $D_1^{c,d}$, the greater $a_{c,d}$, then this implies that $a_{c,d} > a_{c,d,x}$. In other words, the possibility of a catastrophe causing an existential collapse *decreases* WTP to mitigate that catastrophe! Since existential risk is introduced to the model as explicitly welfare decreasing, this conclusion indicates an issue with the model rather than a curious finding about the nature of the welfare effects of changes in existential risk.

This problem is the result of combining two different methodologies to deal with negative utility functions. For the purpose of producing results that reflect the loss in expected future welfare from existential threats, it is satisfactory to

[5] Here I consider a death and consumption catastrophe, though the same results hold for just a death catastrophe as evaluated in Chapter 5.

simply assume that the change in existential threat has the opposite effect. In other words, allow $\delta_{p\prime} = \delta + \delta_p$, rather than $\delta_{p\prime} = \delta - \delta_p$ as suggested in section 5.3. This approach results in $\delta_{p\prime} > \delta$, which would imply that $D_1^{c,d,x} < D_1^{c,d} < D_0$ and yield $a_{c,d,x} > a_{c,d}$ as required. This is the approach taken in computing the results presented in section 5.3.

This approach is sufficient to yield valid results for WTP to mitigate catastrophes that pose a threat to lives, consumption and potentially the entire future of humanity. However, further research is required to unite the endogenous discount model with other models such as that put forward by Martin and Pindyck with a more concrete theoretical grounding.

Appendix B
Alternative Calibrations in Dynamic Willingness to Pay Model

B.1 Recalibrating Pandemic Damages

In section 5.2 I consider a calibration of the expected loss in GDP, given a pandemic occurs, of 2.8% – the value associated with a pandemic killing 14 million people according to modelling in McKibbin and Sidorenko (2006). Here I consider an alternative specification, where given a pandemic occurs, the expected loss to GDP is .7%. Given the possibility of extremely damaging pandemics which could result in significant GDP loss it seems inappropriate to take the lowest calibration for GDP loss, however I apply it here anyway for illustration.

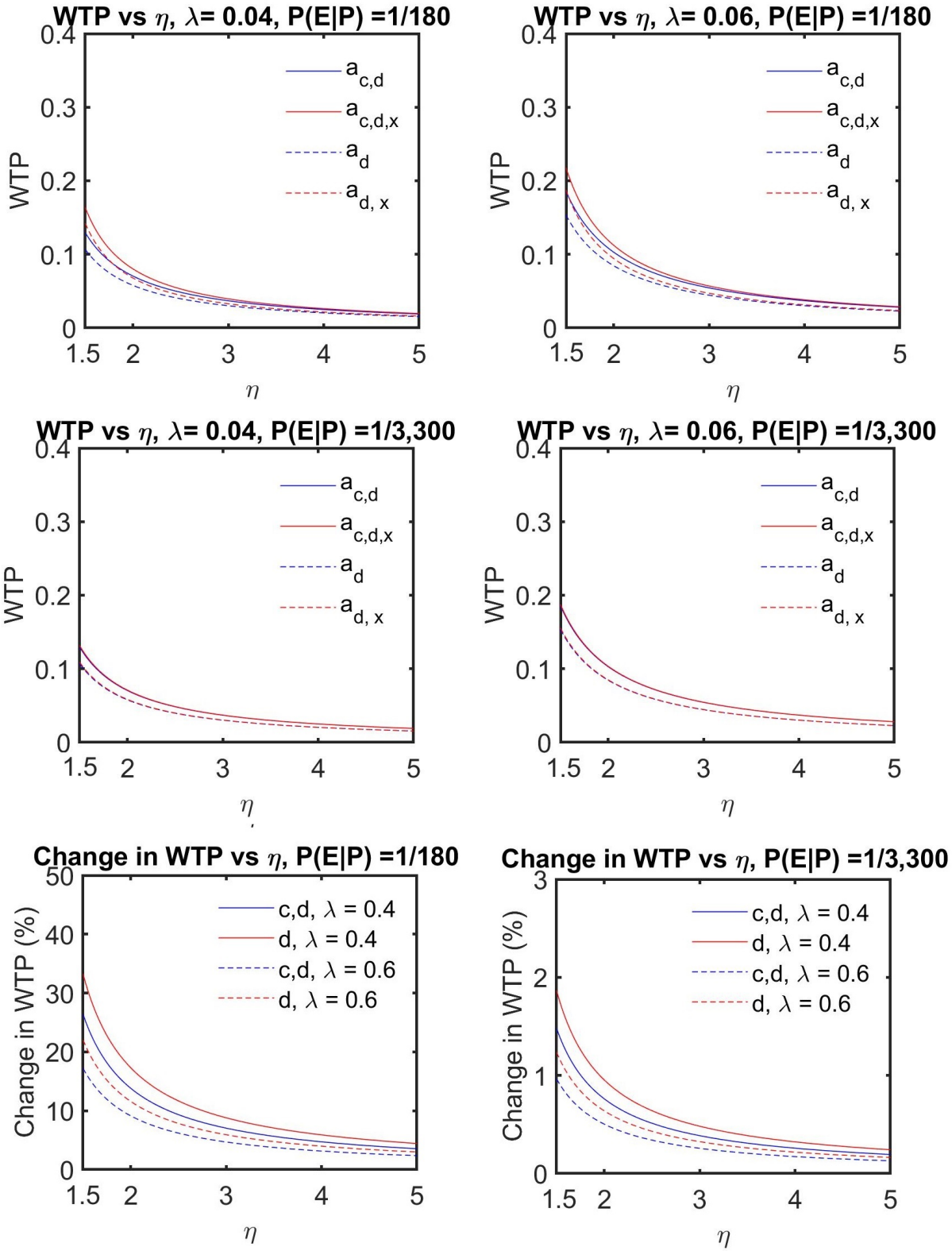
Figure B.1 presents the results of this alternative calibration under no popualtion growth. Relative to Figure 5.1 and Figure 5.2 which represent WTP the 2.8% calibration for expected GDP loss after a pandemic, we see that the WTP under a .7% calibration is significantly less. For example, in the original calibration, when $\lambda = 0.06$ and $P(E|P) = 1/180$, $a_{c,d,x}$ ranges between 25% (at $\eta = 1.5$) and 5% of global GDP (at $\eta = 5$), whereas in the revised calibration this range is between 18% and 2%. However, the percentage change in the WTP as a result of existential considerations presented in Figure B.1 are roughly similar to those in Figure 5.1 and Figure 5.2. As a result, even under different calibrations of the expected economic damages from pandemics, the affect of existential considerations on WTP in proportional terms is relatively constant.

B.2 Recalibrating Value of Statistical Life

Since the results from Chapter 5 are focused on the impact of existential considerations on WTP to mitigate pandemic threats, there was little consideration of underlying parameters used to evaluate WTP that did not interact directly with the (endogenous) discount applied, such as the calibration for the value of a statistical life.

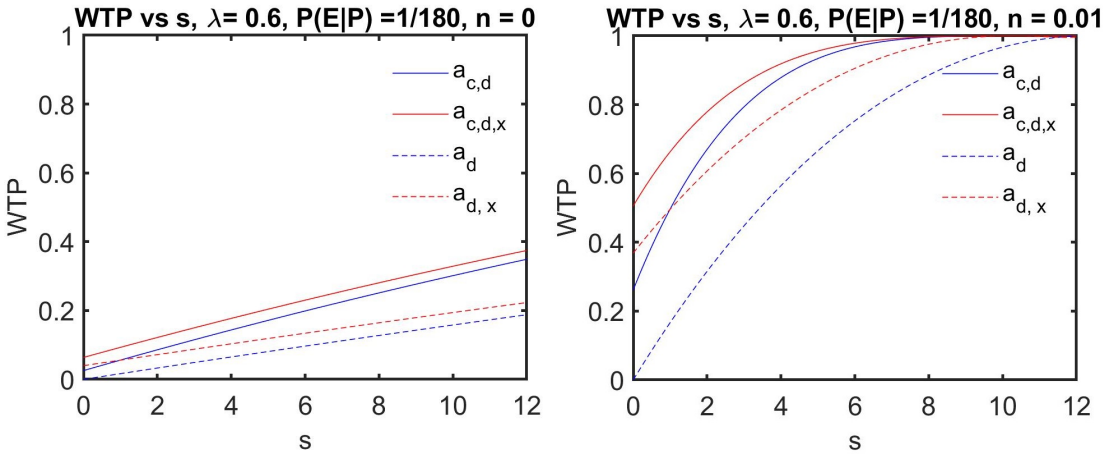
Recalling from Appendix subsection A.7.1, I take $VSL = sw$, where w is lifetime income (or wealth), and s is some positive number chosen by governments

Figure B.1: Willingness to Pay to Avert Mild Pandemics



This figure presents WTP to mitigate all pandemic risk against η when $n = 0$, and the expected GDP loss from a pandemic is .7% (compared to 2.8% applied in Chapter 5). The figure considers two calibrations for the annual probability of pandemic ($\lambda = 0.04$ and $\lambda = 0.06$) and two calibrations for the extinction risk given that a pandemic occurs ($P(E|P) = 1/180$ and $P(E|P) = 1/3,300$).

Figure B.2: Willingness to Pay While Varying s



This figure illustrates how WTP for risk mitigation changes with VSL, where VSL is equal to s multiplied by the average lifetime income. For illustration, this figure calibrates annual pandemic probability, $\lambda = 0.06$, the extinction risk given a pandemic occurs, $P(E|P) = 1/180$, the aversion to intergenerational inequality, $\eta = 1.5$, following preferred the calibration suggested in section 5.4. I also consider two calibrations for population growth, both $n = 0$ and $n = 0.01$, as in section 5.4.

to represent their decision on the value of preventing the death of one of its citizens (ignoring the possibility of their willingness to prevent the death of a non-citizen). Martin and Pindyck (2021) report that this value for s tends to vary quite widely depending on the country and method used to derive this figure,[1] though assume $s = 7$ as a mid-range estimate. Figure B.2 relaxes this assumption and presents WTP while varying s between 0 and 12, under calibrations of $\lambda = 0.06$, $P(E|P) = 1/180$, $\eta = 1.5$ which are the preferred calibrations suggested in section 5.4, and both $n = 0$ and $n = 0.01$. As we can see from the figure, in both $n = 0$ and $n = 0.01$ as s approaches 0, a_d approaches 0 since at $s = 0$ the implication is that there is no value loss associated with a death. However, since the model developed does not apply VSL to evaluate the welfare loss from an existential collapse, we can see that even at low values of s, $a_{d,x}$ is still significantly different from 0.

Comparing the results when $n = 0$ to when $n = 0.01$, we see that with some population growth the WTP for risk mitigation increases very fast as the s increases. This is particularly the case in the calibrations that consider existential risk ($a_{d,x}$ and $a_{c,d,x}$) since under these calibrations there is a higher number of individuals expected to live in the future.

[1] See Viscusi and Aldy (2003).

Appendix C
Sensitivity of Models to Background Risk

To illustrate the WTP for risk mitigation under various assumptions, I have assumed the status quo level of annual extinction risk, δ, to be 0.001 – the level applied in Stern (2007). Since δ here is a subjective estimate, it is unsurprising that there is significant variation in the calibrations for existential risk in the literature – examples of such calibrations are presented in Table C.1.

Table C.1: Calibrations for Existential Risk in the Literature

Source	δ	Corresponding Century Level Risk
Stern (2007)	0.001	9.5 %
Ord (2020)	0.0018	16.7 %
2008 GCR[1] Conference	0.0024	20.0 %
Ng (2016)	≤ 0.0001	≤ 1.0 %
Leslie (1996)	0.0007	6.8 %[2]

This table provides estimates for annual existential risk in notable works on existential risk. Many of these estimates are given over idiosyncratic time periods. For example, the estimate from the 2008 GCR conference was given for 'before 2100', therefore all of these estimates are converted to annual risk using the equation: *Risk before time* $t = \frac{1}{(1+\delta)^t}$.

This variation is a problem for a social planner if these subjective estimates for the background risk have a striking result on the WTP for some risk mitigation activity that has known consequences for the hazard rate. Here I consider an illustrative example to understand how background risk changes the WTP estimate. Suppose a social planner was considering an investment in a project to mitigate risks from Event Y, then the risk endogenous to the social welfare function is δ_Y, while the level of exogenous (background) risk is $\delta_{Y\prime}$, and will determine the value of future

[1] Global Catastrophic Risk.

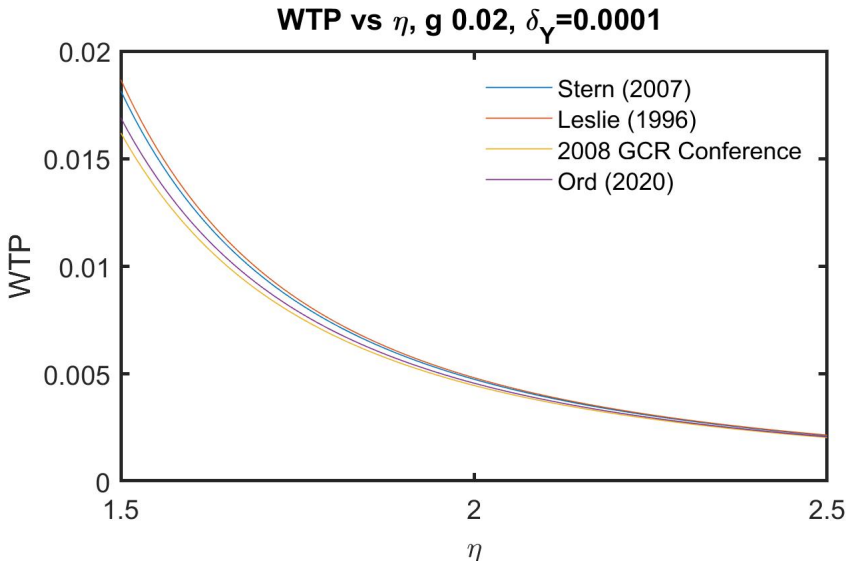
[2] The exact estimate in Leslie (1996) was a 30% existential risk in the next 500 years.

welfare, W_1, if the risk of Event Y is totally mitigated:

$$W_0 = \sum_{t=0}^{\infty} \frac{1}{(1+\delta_Y+\delta_{Y\prime})^t} \sum_{i}^{N} u_{i,t}(c_{i,t})$$
$$W_1 = \sum_{t=0}^{\infty} \frac{1}{(1+\delta_{Y\prime})^t} \sum_{i}^{N} u_{i,t}(c_{i,t})$$

Under the same assumptions as in section 3.1, we can evaluate WTP for a specific project which mitigates all the risk from Event Y, and illustrate the sensitivity of this WTP to changes in background risk, where I let $\delta_Y = 0.0001$ (corresponding to a 1 in 10 000 annual chance of existential collapse form Event Y), and let the total level of risk $(\delta_Y + \delta_{Y\prime})$ correspond to the various estimates for δ in Table C.1.[3] We can see that the sensitivity to the changes in background risk will correspond to how large the endogenous proportion of total risk is, with the results presented in Figure C.1.

Figure C.1: Willingness to Pay for Mitigation of Existential Risk Under Various δ Calibrations



This figure illustrates how different calibrations for the background existential risk change the estimate for WTP to mitigate a fixed amount of existential risk, δ_Y. In this case, I assume that δ_Y makes up 10% of the total existential risk. WTP is considered over the range $1.5 < \eta < 2.5$ since at $\eta > 2.5$ the differences between the various calibrations for background risk is negligible.

As we can see from this figure, the change in background risk only has a modest

[3] I will exclude the 0.0001 calibration suggested by Ng (2016) since it is a clear outlier from other estimates, and it would not make sense in this context unless the social planner thought that Event Y represents all existential risk which I assume is not the case.

impact on the WTP to mitigate all risk of Event Y at low values of η, and at $\eta > 2.5$ different estimates for background risk have a negligible affect on WTP. Therefore, I only consider a single calibration for total risk in each of the analyses presented here corresponding to the Stern (2007) estimate for risk, however one would be able to calibrate the models applied in this thesis to whichever level of background risk they deem appropriate.

Bibliography

P. Allen. The "Justinianic" Plague. *Byzantion*, 49(5):5–20, 1979.

P. Ambrosi, J.-C. Hourcade, S. Hallegatte, F. Lecocq, P. Dumas, and M. H. Duong. Optimal control models and elicitation of attitudes towards climate damages. *Environmental Modeling & Assessment*, 8(3):133–147, 2003. doi: 10.1023/A:1025586922143.

S. C. Antón. Natural history of Homo erectus. *American Journal of Physical Anthropology*, 122(46):126–170, 2003. ISSN 00029483. doi: 10.1002/ajpa.10399.

K. J. Arrow. Global climate change: A challenge to policy. *The Economists' Voice*, 4(3), 2007. doi: 10.2202/1553-3832.1270.

L. Aschenbrenner. Existential Risk and Growth. 2020.

J. C. Avise, D. E. Walker, and G. C. Johns. Speciation durations and pleistocene effects on vertebrate phylogeography. *The Royal Society*, 265:1707–1712, 1998. doi: 10.1142/9789814350709{_}0006.

P. Ball. The lightning-fast quest for COVID vaccines — and what it means for other diseases. *Nature*, 2020. URL https://www.nature.com/articles/d41586-020-03626-1.

O. J. Benedictow. *The Black Death 1346–1353: The Complete History*. The Boydell Press, Woodbridge, 2004.

A. Bommier, B. Lanz, and S. Zuber. Models-as-usual For unusual risks? On the value of catastrophic climate change. *Journal of Environmental Economics and Management*, 74:1–22, 2015. doi: 10.1016/j.jeem.2015.07.003.

N. Bostrom. Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology*, 9, 2002.

N. Bostrom. Existential risk prevention as global priority. *Global Policy*, 4(1):15–31, 2013.

N. Bostrom. The vulnerable world hypothesis. *Global Policy*, 10(4):455–476, 2019.

D. Bricker and J. Ibbitson. *Empty Planet: The Shock of Global Population Decline*. Robinson, London, 2019.

G. Ceballos, P. R. Ehrlich, A. D. Barnosky, A. García, R. M. Pringle, and T. M. Palmer. Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, 1(5):9–13, 2015. ISSN 23752548. doi: 10.1126/sciadv.1400253.

G. Chichilnisky and P. Eisenberger. Asteroids: Assessing catastrophic risks. *Journal of Probability and Statistics*, 2010. doi: 10.1155/2010/954750.

G. Chichilnisky, P. J. Hammond, and N. H. Stern. Fundamental utilitarianism and intergenerational equity with extinction discounting. *Social Choice and Welfare*, 54:397–427, 2020. doi: 10.1007/s00355-019-01236-z.

M. M. Ćirković. Observation and Selection Effects in Global Catastrophic Risks. In N. Bostrom and M. M. Ćirković, editors, *Global Catastrophic Risks*, pages 120–145. Oxford University Press, Oxford, 2008.

H. R. Clarke and W. J. Reed. Consumption/pollution tradeoffs in an environment vulnerable to pollution-related catastrophic collapse. *Journal of Economic Dynamics and Control*, 18(5):991–1010, 1994. ISSN 01651889. doi: 10.1016/0165-1889(94)90042-6.

W. R. Cline. Give greenhouse abatement a fair chance. *International Monetary Fund*, 30(1):56, 1993. doi: 10.5089/9781451953077.022.

M. L. Cropper. Regulating activities with catastrophic environmental effects. *Journal of Environmental Economics and Management*, 3(1):1–15, 1976. ISSN 10960449. doi: 10.1016/0095-0696(76)90009-7.

P. Dasgupta. Commentary: The Stern Review's economics of climate change. *National Institute Economic Review*, 9:4–7, 2007. doi: 10.1177/0027950107077111.

P. Dasgupta. Discounting climate change. *The Journal of Risk and Uncertainty*, 37:141–169, 2008. doi: 10.1007/s11166-008-9049-6.

P. Dasgupta and G. M. Heal. *Economic Theory and Exhaustible Resources*. Cambridge University Press, Cambridge, 1979.

F. S. Dawood, A. D. Iuliano, C. Reed, M. I. Meltzer, D. K. Shay, P. Y. Cheng, D. Bandaranayake, R. F. Breiman, W. A. Brooks, P. Buchy, D. R. Feikin, K. B. Fowler, A. Gordon, N. T. Hien, P. Horby, Q. S. Huang, M. A. Katz, A. Krishnan, R. Lal, J. M. Montgomery, K. Mølbak, R. Pebody, A. M. Presanis, H. Razuri, A. Steens, Y. O. Tinoco, J. Wallinga, H. Yu, S. Vong, J. Bresee, and M. A. Widdowson. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: A modelling study. *The Lancet: Infectious Disease*, 12(9):687–695, 2012. doi: 10.1016/S1473-3099(12)70121-4.

A. P. Dobson, S. L. Pimm, L. Hannah, L. Kaufman, J. A. Ahumada, A. W. Ando, A. Bernstein, J. Busch, P. Daszak, J. Engelmann, M. F. Kinnaird, B. V. Li, T. Loch-Temzelides, T. Lovejoy, K. Nowak, and P. R. Roehrdanz. Ecology and economics for pandemic prevention. *Science*, 369(6502):379–381, 2020. doi: 10.1126/science.abc3189.

M. A. Drupp, M. C. Freeman, B. Groom, and F. Nesje. Discounting disentangled. *American Economic Journal: Economic Policy*, 10(4):109–134, 2018. doi: 10.1257/pol.20160240.

M. Fleurbaey, M. Ferranna, M. Budolfson, F. Dennig, K. Mintz-Woo, R. Socolow, D. Spears, and S. Zuber. The social cost of carbon: Valuing inequality, risk, and population for climate policy. *Monist*, 102(1):84–109, 2019. ISSN 21533601. doi: 10.1093/monist/ony023.

A. Gagnon, M. S. Miller, S. A. Hallman, R. Bourbeau, A. Herring, D. J. Earn, and J. Madrenas. Age-Specific Mortality During the 1918 Influenza Pandemic: Unravelling the Mystery of High Young Adult Mortality. *PLoS One*, 8(8):e69586, 2013. doi: 10.1371/journal.pone.0069586.

J. Galway-Witham and C. Stringer. How did Homo sapiens evolve? *Science*, 360 (6395):1296–1298, 2018. doi: 10.1126/science.aat6659.

R. Garnaut. *The Garnaut Climate Change Review.* Cambridge University Press, Cambridge, 2008.

W. C. Greene. A history of AIDS: Looking back to see ahead. *European Journal of Immunology*, 37:S94–102, 2007. doi: 10.1002/eji.200737441.

H. Greeves. Discounting for public policy: A survey. *Economics & Philosophy*, 33 (3):391–439, 2017. doi: 10.1017/S0266267117000062.

B. Groom and D. Maddison Pr. New Estimates of the Elasticity of Marginal Utility for the UK. *Environmental and Resource Economics*, 72(1):1155–1182, 2019.

R. E. Hall and C. I. Jones. The value of life and the rise in health spending. *The Quarterly Journal of Economics*, 122(1):39–72, 2007. doi: 10.1162/qjec.122.1.39.

K. Harper. Invisible environmental history: Infectious disease in late antiquity. *Late Antique Archaeology*, 12(1):116–131, 2016. doi: 10.1163/22134522-12340069.

M. Harrison. Valuing the Future: The social discount rate in cost-benefit analysis. Technical report, Productivity Commission, Canberra, 2010.

IMF. World Economic Outlook: Managing Divergent Recoveries. Technical report, Washington, DC, 2021.

J. P. A. Ioannidis. Infection fatality rate of COVID-19 inferred from seroprevalence data. *Bulletin of the World Health Organization*, 99(1):19 – 33F, 2020. doi: 10.2471/BLT.20.265892.

C. I. Jones. Why Have Health Expenditures as a Share of GDP Risen So Much? 2002.

K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. Global trends in emerging infectious diseases. *Nature*, 451:990–993, 2008. doi: 10.1038/nature06536.

R. Katz and S. F. Dowell. Revising the international health regulations: Call for a 2017 review conference. *The Lancet Global Health*, 3(7):E352–E353, 2015. doi: 10.1016/S2214-109X(15)00025-X.

E. D. Kilbourne. Influenza pandemics of the 20th century. *Emerging Infectious Diseases*, 12(1):9–14, 2006. doi: 10.3201/eid1201.051254.

L. Klotz. Human error in high-biocontainment labs: A likely pandemic threat, 2019. URL https://thebulletin.org/2019/02/human-error-in-high-biocontainment-labs-a-likely-pandemic-threat/.

A. Leigh. *What's the Worst That Could Happen?: Existential Risk and Extreme Politics*. Random House, New York, 2021.

J. A. Leslie. *The End of the World: the Science and Ethics of Human Extinction*. Routledge, New York, 1996.

W. MacAskill, K. Bykvist, and T. Ord. *Moral Uncertainty*. Oxford University Press, Oxford, 2020.

I. W. R. Martin and R. S. Pindyck. Welfare costs of catastrophes: Lost consumption and lost lives. *The Economic Journal*, 131(634):946–969, 2021. doi: 10.1093/ej/ueaa099.

W. J. McKibbin and A. A. Sidorenko. Global macroeconomic consequences of pandemic influenza. Technical report, The Crawford School of Public Policy, 2006. URL https://cama.crawford.anu.edu.au/pdf/working-papers/2006/262006.pdf.

A. Méjean, A. Pottier, S. Zuber, and M. Fleurbaey. Catastrophic climate change, population ethics and intergenerational equity. *Climatic Change*, 163:873–890, 2020. doi: 10.1007/s10584-020-02899-9.

P. Millet and A. Snyder-Beattie. Existential risk and cost-effective biosecurity. *Health Security*, 15(4):373–383, 2017. doi: 10.1089/hs.2017.0028.

A. Millner. On welfare frameworks and catastrophic climate risks. *Journal of Environmental Management*, 65:310–325, 2013. doi: 10.1016/j.jeem.2012.09.006.

J. A. Mirrlees. Optimum growth when technology is changing. *The Review of Economic Studies*, 34(1):95–124, 1967. doi: 10.2307/2296573.

L. Mordechai and M. Eisenburg. Rejecting catastrophe: The case of the Justinianic Plague. *Past and Present*, 244(1):3–50, 2019. doi: 10.1093/pastj/gtz009.

L. Mordechai, M. Eisenberg, T. P. Newfield, A. Izdebski, J. E. Kay, and H. Poinar. The Justinianic Plague: An inconsequential pandemic?. *Proceedings of the National Academy of Sciences of the United States of America*, 116(51):25546–25554, 2019. doi: 10.1073/pnas.1903797116.

National Research Council. *Globalization, Biosecurity, and the Future of the Life Sciences*. The National Academies Press, Washington, DC, 2006. doi: 10.17226/11567.

R. G. Newell, W. A. Pizer, and B. C. Prest. A Discounting Rule for the Social Cost of Carbon. 2021.

Y.-K. Ng. Social criteria for evaluating population change: An alternative to the Blackorby-Donaldson criterion. *Journal of Public Economics*, 29(3):375–381, 1986. doi: 10.1016/0047-2727(86)90036-8.

Y.-K. Ng. The importance of global extinction in climate change policy. *Global Policy*, 7(3):315–322, 2016. doi: 10.1111/1758-5899.12318.

W. D. Nordhaus. The 'Stern Review' on the Economics of Climate Change. 2006.

W. D. Nordhaus. *A Question of Balance*. Yale University Press, New Haven, 2008.

W. D. Nordhaus. The economics of tail events with an application to climate change. *Review of Environmental Economics and Policy*, 5(2):240–257, 2011. doi: 10.1093/reep/rer004.

A. Nouri and C. F. Chyba. Biotechnology and biosecurity. In N. Bostrom and M. M. Ćirković, editors, *Global Catastrophic Risks*, pages 450–480. Oxford University Press, Oxford, 2008.

T. Ord. *The Precipice: Existential Risk and the Future of Humanity*. Bloomsbury Publishing, London, 2020.

J. Z. Oreskes, C. N. W. Naomi, M. Williams, A. D. Barnosky, A. Cearreta, P. Crutzen, E. Ellis, M. A. Ellis, I. J. Fairchild, J. Grinevald, P. K. Haff, I. Hajdas, R. Leinfelder, J. McNeill, E. O. Odada, C. Poirier, D. Richter, W. Steffen, C. Summerhayes, J. P. Syvitski, D. Vidas, M. Wagreich, S. L. Wing, A. P. Wolfe, Z. An, and N. Oreskes. When did the Anthropocene begin? A mid-twentieth century boundary level is stratigraphically optimal. *Quaternary International*, 383:196–203, 2015. doi: 10.1016/j.quaint.2014.11.045.

D. Parfit. *Reasons and Persons*. Oxford University Press, Oxford, 1984.

R. A. Posner. *Catastrophe: Risk and Response*. Oxford University Press, Oxford, 2004.

F. P. Ramsey. A mathematical theory of saving. *The Economic Journal*, 38(152): 543–559, 1928.

W. T. Reid. Properties of solutions of an infinite system of ordinary linear differential equations of the first order with auxiliary boundary conditions. *Transactions of the American Mathematical Society*, 32(2):284–318, 1930.

M. Roser, H. Ritchie, and E. Ortiz-Ospina. World Population Growth. *Our World in Data*, 2013. URL https://ourworldindata.org/world-population-growth.

M. Roser, H. Ritchie, E. Ortiz-Ospina, and J. Hasell. Coronavirus Pandemic (COVID-19). *Our World in Data*, 2020. URL https://ourworldindata.org/coronavirus.

A. Sandberg and N. Bostrom. Global catastrophic risk survey. 2008.

F. Sanmarchi, D. Golinelli, J. Lenzi, F. Esposito, A. Capodici, C. Reno, and D. Gibertoni. Exploring the gap between excess mortality and COVID-19 deaths in 67 countries. 4(7):1–5, 2021. doi: 10.1001/jamanetworkopen.2021.17359.

P. Schulte, L. Alegret, I. Arenillas, J. A. Arz, P. J. Barton, P. R. Bown, T. J. Bralower, G. L. Christeson, P. Claeys, C. S. Cockell, G. S. Collins, A. Deutsch, T. J. Goldin, K. Goto, J. M. Grajales-Nishimura, R. A. Grieve, S. P. Gulick, K. R. Johnson, W. Kiessling, C. Koeberl, D. A. Kring, K. G. MacLeod, T. Matsui, J. Melosh, A. Montanari, J. V. Morgan, C. R. Neal, D. J. Nichols, R. D. Norris, E. Pierazzo, G. Ravizza, M. Rebolledo-Vieyra, W. U. Reimold, E. Robin, T. Salge, R. P. Speijer, A. R. Sweet, J. Urrutia-Fucugauchi, V. Vajda, M. T. Whalen, and P. S. Willumsen. The chicxulub asteroid impact and mass extinction at the cretaceous-paleogene boundary. *Science*, 327(5970):1214–1218, 2010. ISSN 10959203. doi: 10.1126/science.1177265.

P. M. Sharp and B. H. Hahn. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*, 1(1):a006841, 2011. doi: 10.1101/cshperspect.a006841.

N. H. Stern. *The Stern Review: The Economics of Climate Change*. Cambridge University Press, Cambridge, 2007.

N. Taleb, Y. Bar-Yam, R. Douady, J. Norman, and R. Read. The Precautionary Principle: Fragility and Black Swans from Policy Actions. 2014a.

N. N. Taleb, R. Read, R. Douady, J. Norman, and Y. Bar-Yam. The Precautionary Principle (with Application to the Genetic Modification of Organisms). 2014b. URL http://arxiv.org/abs/1410.5787.

J. K. Taubenberger and D. M. Morens. 1918 Influenza: The Mother of All Pandemics. *Revista Biomedica*, 17(1):69–79, 2006.

The Centre for Long-Term Resilience. Future Proof: The Opportunity to Transform the UK's Resilience to Extreme Risks. Technical report, 2021.

J. B. Tucker and R. A. Zilinskas. The promise and perils of synthetic biology. *The New Atlantis*, 12(1):25–45, 2006.

United Nations. World Population Prospects 2019. Technical report, United Nations Population Division, 2019. URL http://population.un.org/wpp/.

United Nations. Global HIV & AIDS statistics — Fact sheet. Technical report, United Nations, 2020.

United Nations Population Division. World Population Prospects: The 2019 Revision. Technical report, 2019. URL https://population.un.org/wpp2019/Download/Standard/Interpolated/.

US Department of Defence. Narrative Summaries of Accidents Involving Nuclear Weapons 1950-1980. Technical report, Homeland Security Digital Library, 1981.

C. Viboud, L. Simonsen, R. Fuentes, J. Flores, M. A. Miller, and G. Chowell. Global mortality impact of the 1957-1959 influenza pandemic. *Journal of Infectious Disease*, 1(213):738–745, 2016. doi: 10.1093/infdis/jiv534.

W. K. Viscusi and J. E. Aldy. The value of a statistical life: A critical review of market estimates throughout the world. *The Journal of Risk and Uncertainty*, 27 (1):5–76, 2003. doi: 10.1023/A:1025598106257.

C. R. Watson, M. Watson, D. Gastfriend, and T. K. Sell. Federal funding for health security in FY2019. *Health Security*, 16(5):281–303, 2018. doi: 10.1089/hs.2018.0077.

M. L. Weitzman. A review of 'The Stern Review on the Economics of Climate Change'. *The Journal of Economic Literature*, 45(3):703–724, 2007.

M. L. Weitzman. On modelling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 91(1):1–19, 2009.

G. Wilson. Minimizing global catastrophic and existential risks from emerging technologies through international law. *Virginia Environmental Law Journal*, 31 (2):307–364, 2013.

World Bank. People, Pathogens, and Our Planet: The Economics of One Health. Technical report, World Bank, Washington, DC, 2012. URL https://openknowledge.worldbank.org/handle/10986/11892.

World Health Organization. World Health Assembly Resolution 58.3. Technical report, World Health Organization, Geneva, 2005. URL https://www.who.int/ipcs/publications/wha/ihr_resolution.pdf.

P. Ziegler. *The Black Death*. John Day Co., New York, 1969.

S. Zuber, N. Venkatesh, T. Tännsjö, C. Tarsney, H. O. Stefánsson, K. Steele, D. Spears, J. Sebo, M. Pivato, T. Ord, Y. K. Ng, M. Masny, W. Macaskill, N. Lawson, K. Kuruc, M. Hutchinson, J. E. Gustafsson, H. Greaves, L. Forsberg, M. Fleurbaey, D. Coffey, S. Cato, C. Castro, T. Campbell, M. Budolfson, J. Broome, A. Berger, N. Beckstead, and G. B. Asheim. What should we agree on about the repugnant conclusion? *Utilitas*, pages 1–5, 2021. ISSN 17416183. doi: 10.1017/S095382082100011X.